

**TEMPORAL PREDICTIVE MODEL FOR FORT HOOD, TEXAS:
A PILOT STUDY IN THE COWHOUSE CREEK DRAINAGE**

by

Joshua S. Campbell

William C. Johnson, Ph.D.

*Department of Geography
University of Kansas*

**United States Army Fort Hood
Archaeological Resource Management Series
Research Report No. 54**

and

**Center for Environmental Management of Military Lands
Technical Publication Series
TPS 04-10**

June 2004

**TEMPORAL PREDICTIVE MODEL FOR FORT HOOD, TEXAS:
A PILOT STUDY IN THE COWHOUSE CREEK DRAINAGE**

Prepared for

Directorate of Public Works
Environmental Division
Fort Hood

by

Center for Environmental Management of Military Lands
Colorado State University
Fort Collins, CO 80523-1490

In partial fulfillment of
Agreement No. DAMD17-02-2-0008
Task Order 2002-4

June 2004

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) June 2004		2. REPORT TYPE Final Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Temporal Predictive Model for Fort Hood, Texas: A Pilot Study in the Cowhouse Creek Drainage.				5a. CONTRACT NUMBER Agreement No. DAMD17-02-2-0008	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Joshua S. Campbell and William C. Johnson				5d. PROJECT NUMBER	
				5e. TASK NUMBER Task Order 2002-2	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Environmental Management of Military Lands Colorado State University, Campus Delivery 1490 Fort Collins, Colorado 80523-1490				8. PERFORMING ORGANIZATION REPORT NUMBER CEMML TPS 04-10	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Fort Hood Directorate of Public Works Environmental Division 111 Corps and Fort Hood Fort Hood, Texas 76544				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) ARMS Series, Research Report No. 54	
12. DISTRIBUTION/AVAILABILITY STATEMENT Available for public release					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This project represents a pilot study in the Cowhouse Creek drainage of Fort Hood, Texas, with the aim of creating an interactive archaeological predictive modeling tool to assist decision making on the part of the Fort Hood CRM program. It resulted in three GIS data products: 1) a statistically based two-dimensional (2-D) surface probability archaeological predictive model, 2) an updated alluvial landform map for the study area used in a 3-D sub-surface probability predictive model and 3) an interactive "Predictive Model Viewer" tool that integrates the 2-D and 3-D models. The unique methodological approach used here results in the integration of surface and sub-surface probability models and the interactive tool provides a rapid, easy-to-use method of synthesizing their output for management purposes.					
15. SUBJECT TERMS Archaeology, Fort Hood, Texas, prehistory, historic preservation, predictive archaeological modeling, Cowhouse Creek drainage, integrated 2-D and 3-D predictive model, interactive "Predictive Model Viewer" tool					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unclassified	18. NUMBER OF PAGES 60	19a. NAME OF RESPONSIBLE PERSON Dr. James A. Zeidler
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) 970-491-3741

TABLE OF CONTENTS

LIST OF FIGURES.....	ix
ABSTRACT	xi
ACKNOWLEDGEMENTS	xii
PREFACE	xiii
1.0 Introduction.....	1
2.0 Study Area and Landscape Characterization.....	1
3.0 Archaeological (2-D) Predictive Modeling Review.....	3
3.1 Fundamental Components of Predictive Models.....	5
3.1.1 Unit of Investigation	5
3.1.2 Archaeological Events	6
3.1.3 Predictive Models as Decision Rules.....	6
3.2 Factors Influencing Model Development	7
3.3 Limitations	8
3.3.1 Garbage In – Garbage Out	8
3.3.2 Environmental Determinism.....	8
3.4 Use of Geographic Information Systems	9
3.5 Logistic Regression	9
4.0 Methodology	9
4.1 Database Construction	9
4.2 Archaeological Database.....	10
4.2.1 Site Data.....	10
4.2.2 Non-Site Data.....	10
4.3 Environmental Database.....	11
4.3.1 High Resolution Terrain Dataset	11
4.3.2 Hydrographic Dataset.....	12
4.3.3 Geomorphic Dataset.....	13
5.0 Model Development and Results.....	14
5.1 Statistical (2-D) Model.....	15
5.1.1 Model Output	16
5.1.2 Model Assessment.....	16
5.1.3 2-D Model Conclusions	27
5.2 Geoarchaeological (3-D) Model.....	28
6.0 Model Integration	28
7.0 Conclusions.....	32
7.1 Expanding the Methodology.....	32
REFERENCES CITED	33
APPENDIX A: GIS DATA LAYERS	35
APPENDIX B: PREDICTIVE MODEL VIEWER (NOT AVAILABLE TO THE GENERAL PUBLIC).....	45
ENCLOSED CD-ROM	Inside Back Cover
Electronic manuscript in .PDF format including all color figures	

LIST OF FIGURES

Figure 1. Major Land Resource Areas in Texas	2
Figure 2. Fort Hood Military Reservation Boundary and Study Area.	3
Figure 3. Digital Elevation Model with Study Area Boundary.	4
Figure 4. Study area 3-D perspective view using high resolution imagery draped over a DEM...	5
Figure 5. NRHP Eligible Known Site Training Sample	11
Figure 6. Randomly Generated Non-Site Training Sample.	12
Figure 7. Digital Elevation Model with Hydrology.	13
Figure 8. Alluvial Landforms of Cowhouse Creek and Table Rock Creek.....	14
Figure 9. Probability Distribution Output of Predictive Model.	17
Figure 10. Predictive Probability of NRHP-Eligible Site Training Sample.	19
Figure 11. Predictive Probability of Known Non-Sites Training Sample.....	19
Figure 12. Optimal Cut-Point of Predictive Model (training data).....	20
Figure 13. NRHP Non-Eligible Known Site Testing Sample.	21
Figure 14. Randomly Generated Non-Site Testing Sample.	22
Figure 15. Predicted Probability of NRHP Non-Eligible Testing Sample.	23
Figure 16. Predicted Probability of Known Non-Site Testing Sample.....	23
Figure 17. Optimal Cut-Point of Predictive Model (testing data).....	24
Figure 18. Performance Curves.	25
Figure 19. Percent Gain Over Random Classification.....	25
Figure 20. High Probability Areas at the 0.25 Cut-point.	26
Figure 21. Screen shot of Predictive Model Viewer Graphical User Interface (GUI).	30
Figure 22. Screen shot of Predictive Model Viewer in operation.	30
Figure 23a. Cross-section of Cowhouse Creek drainage.	31
Figure 23b. Cross-section of Table Rock Creek drainage.....	31
Figure 24. Distribution of Prehistoric Archaeological Sites.	37

Figure 25. Derived Slope.....	38
Figure 26. Relief within 150 meters.	39
Figure 27. Relief within 300 meters	40
Figure 28. Shelter Index.	41
Figure 29. Distance to Primary Streams.....	42
Figure 30. Distance to Secondary Streams.	43

LIST OF TABLES

Table 1. Accuracy assessment using a 0.25 cut point	20
Table 2. Results of the testing sample at the 0.25 cut-point level	24
Table 3. Preservation potentials for surface and subsurface cultural sites (- none; + low; ++ medium; +++ high; + or ++ facies dependent)	29

ABSTRACT

This project represents a pilot study in the Cowhouse Creek drainage of Fort Hood, Texas, with the aim of creating an interactive archaeological predictive modeling tool to assist decision making on the part of the Fort Hood CRM program. It resulted in three GIS data products: 1) a statistically based two-dimensional (2-D) surface probability archaeological predictive model; 2) an updated alluvial landform map for the study area used in a 3-D sub-surface probability predictive model; and 3) an interactive “Predictive Model Viewer” tool that integrates the 2-D and 3-D models. The unique methodological approach used here results in the integration of surface and sub-surface probability models and the interactive tool provides a rapid, easy-to-use method of synthesizing their output for management purposes.

ACKNOWLEDGEMENTS

The authors express their gratitude to the Fort Hood Cultural Resources Program, specifically Dr. Cheryl Huckerby (Program Manager), Karl Kleinbach, Kristen Wenzel, and Dennis Glinn, for their assistance in providing GIS datasets and access to archaeological sites.

PREFACE

This report is part of a project executed by the Center for Environmental Management of Military Lands (CEMML), Colorado State University, on behalf of the Fort Hood Cultural Resources Management Program. The project was entitled “Temporal Predictive Archaeological Model for Fort Hood, Texas,” and was contracted through the United States Army Medical Research Acquisition Activity (USAMRAA), Fort Detrick, Maryland, under the terms of Cooperative Agreement DAMD17-02-2-0008, Task Order 2002-2, awarded to CEMML. Project execution involved a period of fieldwork during which NRHP-eligible prehistoric sites within the Cowhouse Creek drainage were re-located and georeferenced using state-of-the-art GPS technology with decimeter accuracy. This permitted precise verification and/or correction of site locations prior to the predictive modeling analysis. This work was carried out by Mr. Stephen A. Sherman, CEMML staff archaeologist, with the assistance of Mr. Joel Gutierrez. In addition, CEMML GIS Analyst Michael O’Donnell created the Digital Elevation Model (DEM) employed in the predictive modeling analysis, as well as a series of other GIS datasets.

The 2-D and 3-D predictive modeling analysis contained in this report was sub-contracted to Professor William C. Johnson, Department of Geography, University of Kansas, through Subaward Agreement G-2496-1 between Colorado State University and the University of Kansas.

James A. Zeidler, Ph.D., RPA
Principal Investigator

1.0 INTRODUCTION

Project requirements, as outlined by the Statement of Work for the pilot three-dimensional (3-D) predictive model of Fort Hood, Texas, involve creation of two products: first, an *explicit GIS layer and associated database* usable as a predictive model for determining where additional unknown sub-surface cultural materials may be located within the pilot study area; and second, a *methodology* that the Fort Hood CRM personnel can use to expand the study to other priority areas on Fort Hood. In order to fulfill the project requirements, three Geographic Information Systems (GIS) data products were created: 1) a two-dimensional (2-D) archaeological predictive model; 2) an updated alluvial landform map for upper Cowhouse Creek and Table Rock Creek, which forms the basis of the 3-D predictive model; and 3) the 'Predictive Model Tool', a custom-software tool designed to integrate the results of the 2-D and 3-D models.

Specific report objectives include: 1) organize a GIS database containing archaeological and environmental data in an appropriate format for predictive modeling; 2) construct a 2-D archaeological predictive model using quantitative landscape metrics; 3) translate the existing geoarchaeological (3-D) model into a GIS data layer (Nordt 1992); and 4) integrate the output of the two models using a customized software tool. The resulting data products provide a rapid method for investigating the cultural potential of any given land parcel within the study area. Using this methodology, a 2-D and 3-D integrated model can be constructed for the entire Fort Hood Military Reservation. Software used in the construction of these models includes ESRI ArcGIS and Spatial Analyst GIS tools, Statistical Package for the Social Sciences (SPSS) statistical software, Microsoft Excel spreadsheet, and Microsoft Visual Basic.

Two different types of predictive models are used in this report: first, a statistically based surface probability model, referred to as the 2-dimensional (2-D) model; and second, a geoarchaeological surface and subsurface model. The 2-D model analyzes the spatial pattern observed in a sample of archaeological sites and quantifies the relationship between site locations and the local environment. Empirical patterns observed in the sample data are

then projected onto the study area and tested for accuracy against a different set of archaeological sites. The 2-D model does not take into consideration the possibility of sites buried at significant depth; it is designed to predict areas of the landscape with similar environmental conditions, on the surface, to areas with known cultural resources. The 3-D model is based on an investigation of the extent and composition of alluvial landforms in the Cowhouse Creek and Table Rock Creek valleys. A generalized composite geologic cross-section for each river valley was previously created that describes the potential stratigraphic composition of individual landforms and the implications for cultural preservation in those landforms, both on the surface and buried. A custom-built software tool, the 'Predictive Model Viewer', is designed to integrate the output of these two models in a concise form within the ArcGIS environment.

Fort Hood archaeological and environmental GIS datasets are well suited for the development of archaeological predictive models. Methods used in the integrated model represent a unique approach for combining high-resolution environmental data, an intensive archaeological survey regime, and an existing well-developed geoarchaeological model into a cohesive model of the potential for cultural material within the study area. These facets of the Fort Hood dataset, along with the increased usability of GIS, combine to result in a data product superior to previous modeling attempts. The integrated GIS model allows a user to select any location within the study area and instantly receive model predictions concerning the probability of that location for containing both a surface and/or buried archaeological site and the cultural/temporal period of the potential site.

2.0 STUDY AREA AND LANDSCAPE CHARACTERIZATION

Fort Hood is located in a dissected part of the eastern margin of the Edwards Plateau of central Texas and consists of Lower Cretaceous sedimentary geology. United States Department of Agriculture data on the Major Land Resource Areas indicates Fort Hood is located in the Grand Prairie physiographic zone (Figure 1). Stratigraphy of the military reservation and environs consists of the Glen Rose Forma-

tion (limestone), Paluxy Sand (quartz sands), Walnut Clay (clay, limestone and shale), Comanche Peak Limestone, Edwards Limestone and Kiamichi Clay (undifferentiated) (Barnes 1979). Earliest post-Cretaceous landscape evolution still in evidence is the Lampasas Cut Plain on the Edwards Plateau (Hayward et al. 1990) (Manning Surface: Nordt 1992). Subsequently, an intermediate surface (Killeen Surface: Nordt 1992) developed through pedimentation, resulting in a rolling plain developed on the Walnut Clay. Middle Pleistocene abandonment of the Killeen Surface was followed by an episode of drainage entrenchment (Hayward et al. 1990), which resulted in a sequence of late-Pleistocene and Holocene terraces and the modern flood plain (Nordt 1992).

Late-Quaternary alluvial and environmental history of the region has been well articulated by Blum and colleagues. In a study of the Perdarnales River system of central Texas, Blum and Valastro (1989) reconstructed the alluvial history, focusing on detailed articulation of the late-Holocene record. Blum (1990), in an investigation of the Pleistocene and Holocene alluvial history of the Colorado River system of central Texas, considered climatic and eustatic impacts as interpreted from the alluvial stratigraphic record. Employing existing fossil vertebrate, pollen, and plant macrofossil data from the Edwards Plateau as climatic proxies, Toomey and others (1993) reconstructed a temporal sequence of change in temperature and effective moisture from about 14,000 years ago. Blum and others (1994) developed a model of

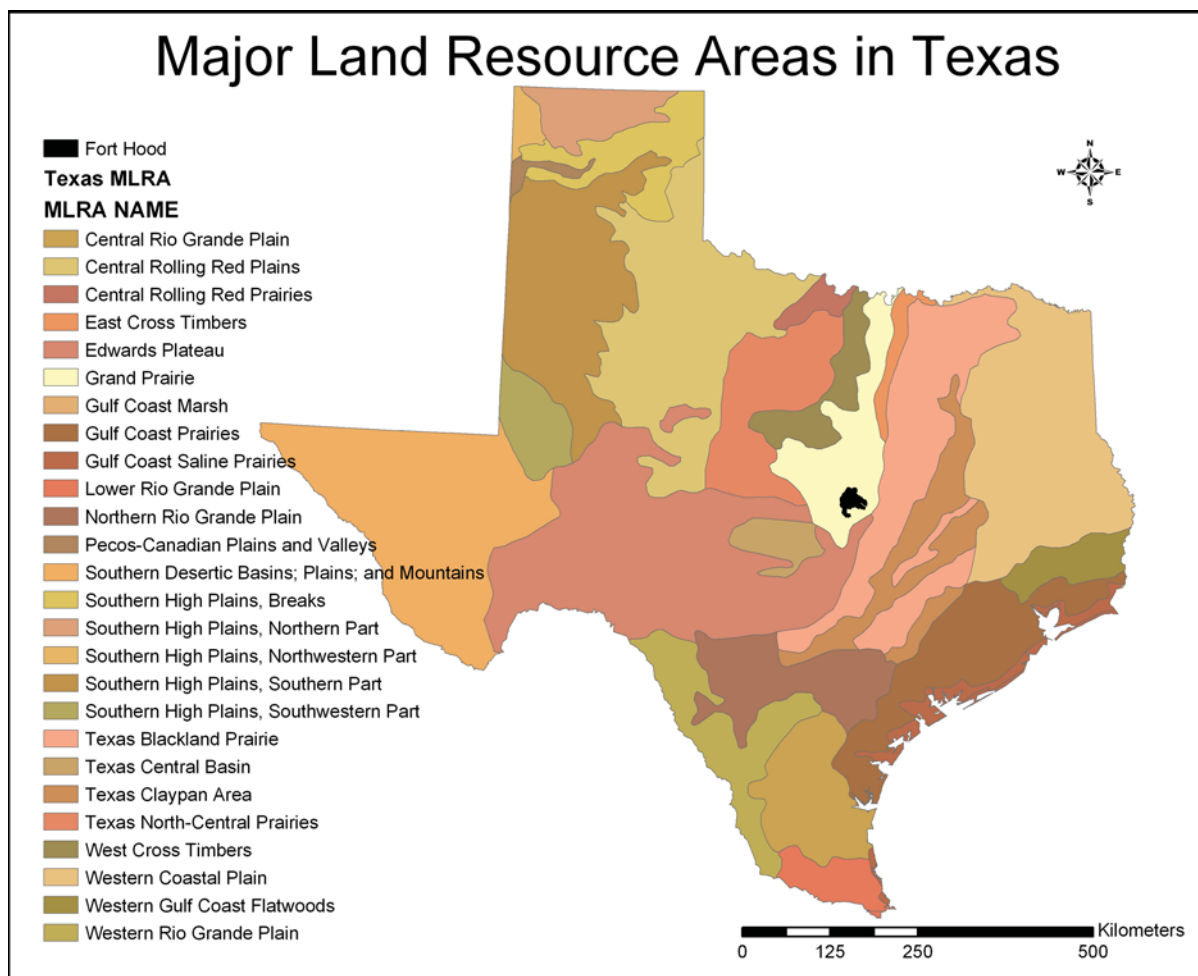


Figure 1. Major Land Resource Areas in Texas

fluvial landscape evolution in the stream systems draining the Edwards Plateau, spanning the last 20,000 years (Barnes 1979; Blum 1990; Blum et al. 1994; Blum and Salvatore Valastro 1989; Hayward et al. 1990; Toomey et al. 1993).

Specifically, the study area of this project is located in the western portion of the Fort Hood Military Reservation and includes Upper Cowhouse Creek and Table Rock Creek (Figure 2 and Figure 3). The western boundary of the study area is Highway 116 and the eastern boundary is West Range Road. The northern and southern boundaries are demarcated by Shell Mountain and Manning Mountain Roads on the north, and by Elijah Road on the south. The area contains components of the stream networks and the intermediate upland, Killeen

Surface. Figure 4 displays the study area as rendered in 3-D using ESRI ArcScene software. Created by draping high-resolution aerial photography over the existing DEM, the image perspective is looking northwest, up the Cowhouse Creek valley.

3.0 ARCHAEOLOGICAL (2-D) PREDICTIVE MODELING REVIEW

This section provides an overview of the theoretical background of archaeological predictive modeling. In essence, an archaeological predictive model is a tool that indicates the likelihood of cultural material being present at a location (Gibbon 2000). The theoretical basis of predictive modeling relies on human behavior being non-random, and that the spatial pat-

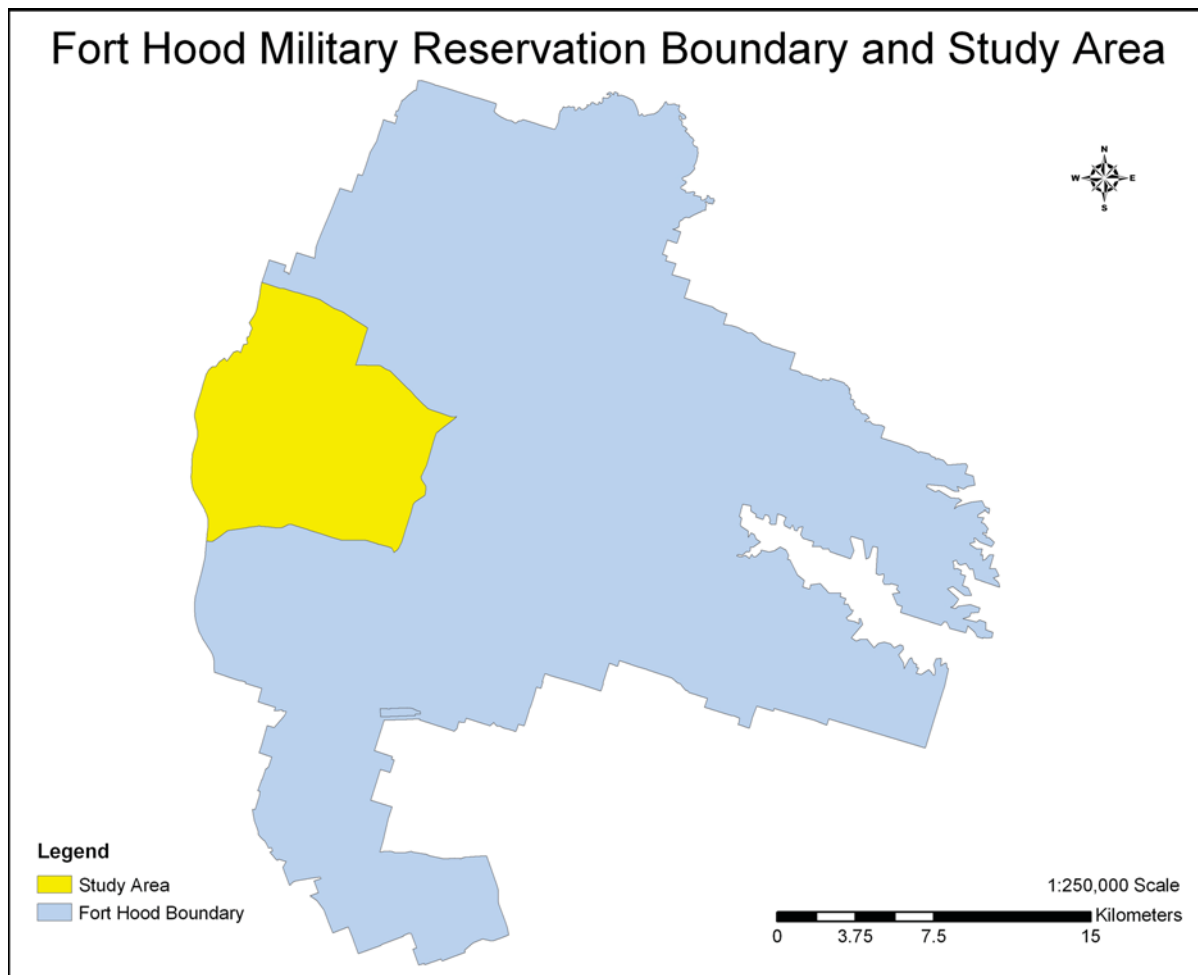


Figure 2. Fort Hood Military Reservation Boundary and Study Area.

tern of cultural materials on the landscape represents the remnants of an intentional strategy to exploit landscape resources. Predictive models attempt to extract the spatial pattern inherent to a sample of site locations with respect to various environmental variables (using any number of pattern recognition methods) and project the abstracted pattern to the study area as a whole (Kvamme 1992). By applying the quantitative abstraction to the entire study area, the model selects locations with a set of landscape characteristics similar to those of the input sample of known site locations. Identifying these areas of the landscape should increase the likelihood of finding unknown sites.

The majority of archaeological models developed for North America use environmental data to determine the probability of a loca-

tion for containing cultural material. Predictive models, when combined with GIS, can be thought of as macro-scale landscape screening tools. The prediction, or screening, component of these models becomes apparent when the empirical relationship extracted from the sample data is projected onto areas not surveyed for archaeological sites. GIS allows the compilation of environmental datasets covering very large tracts of land; with the development of an equation relating environmental conditions to site locations, these large unsurveyed areas can be 'screened' for the potential of containing sites. In terms of cultural resource management or development planning, having information about the potential location of sites can save time and money (Hudak et al. 2000).

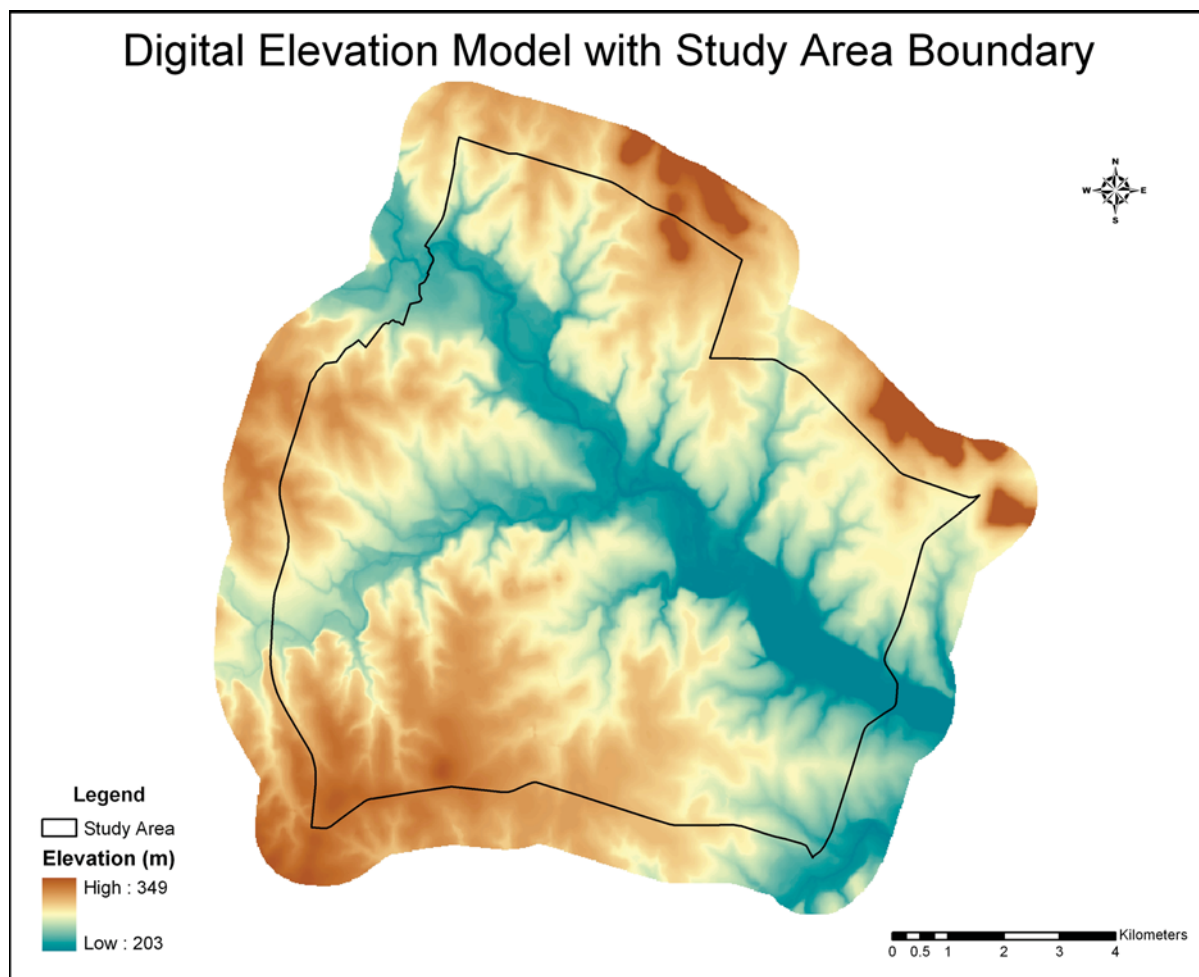


Figure 3. Digital Elevation Model with Study Area Boundary.

A more specific explanation of archaeological predictive models is offered by Kvamme (1990:261), who defines a predictive model as “an *assignment procedure* that correctly indicates an *archaeological event* outcome at a *land parcel* location with greater *probability than that attributable to chance*.” The assignment procedure, or decision rule, is a set of criteria that determine whether a land parcel is assigned to one archaeological event class or another on the basis of some non-archaeological input. In the case of environmentally based archaeological predictive modeling, the decision rule uses environmental information about a land parcel as input. Output of the decision rule is classification of the land parcel to a single archaeological event class (Kvamme 1990). The following section discusses the ideas of a land parcel, archaeological event classes, and decision rules in greater detail.

3.1 Fundamental Components of Predictive Models

3.1.1 Unit of Investigation

The fundamental component of any archaeological predictive model is the unit of investigation. Typically, in archaeological studies the analysis unit is the archaeological site, but, in the case of archaeological predictive modeling the unit of investigation is the individual parcel of land (Kvamme 1988a). Dividing the landscape into a series of contiguous parcels is analogous to laying a grid of uniform size over the landscape. All land parcels therefore will be uniform in size and shape. This division works well with the use of GIS, as the single land parcel forms the standard grid cell used in raster data analysis. Determining the appropriate parcel size involves consideration of the available environmental data and the modeling goals. Size of the land parcel has implications for the conclusions of the model. If the parcel size is large ($> 1\text{km}^2$), the results may be too coarse

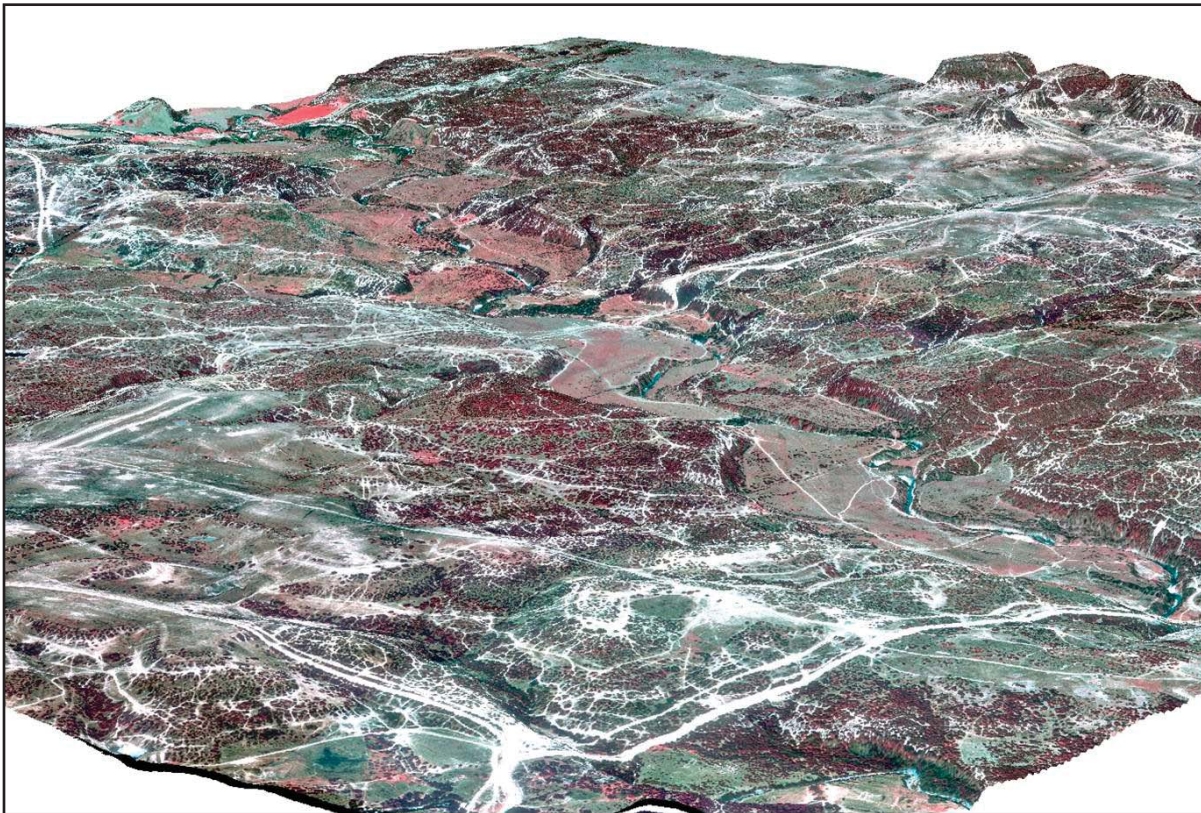


Figure 4. Study area 3-D perspective view using high resolution imagery draped over a DEM

for any actual site predictions. However, if the goal is to predict the number of sites that occur within one land parcel, a large parcel size may be desirable. The optimal parcel size captures the variability of the landscape that influenced cultural behavior but is not a finer scale than the available environmental data (Hudak et al. 2000). Consideration of the available environmental datasets is important because although most GIS packages will display map data at any scale, digital datasets are collected with a specific margin of error and consequently have limits to the positional accuracy of the data (Clarke et al. 2002). Therefore, the use of a grid size that is too small for the mapping scale of the environmental data runs the risk of introducing false precision into the model.

At Fort Hood, the methods used to derive the available terrain data utilized modern photogrammetric techniques and as a result, the data are considered accurate and precise. Subsequently, the terrain dataset is very high resolution, and a small land parcel size (5 meters x 5 meters) is possible. Quality of the available environmental data is discussed later in the report, however, it should be noted the current 2-D model utilizes a land parcel size significantly smaller than most predictive models cited in the literature.

3.1.2 Archaeological Events

Output of an archaeological predictive model is the assignment of a land parcel to an archaeological event class. Prior to model construction, the archaeological event classes must be defined. The simplest set of archaeological events involves classifying a parcel into either a *site-present* or *site-absent* class. Other predictive models use archaeological event classes structured to predict the type of site present at a location, the number of sites within a parcel, or the density of artifacts within a parcel. Regardless of the modeling goals, the set of potential event classes must be mutually exclusive and exhaustive, meaning a parcel must be assigned to only one of the event classes and all parcels must be classified (Kvamme 1990).

Using notation derived from Kvamme (1990), the following sections describe the potential event classes available to the current 2-D predictive model. For each land parcel used to construct the model, two potential archaeo-

logical events representative of the land parcels true condition are possible:

$$\begin{aligned} S &= \{\text{site-present}\} \\ &\text{or} \\ S' &= \{\text{site-absent}\} \end{aligned}$$

Model output assigns every land parcel into one of two potential archaeological event classes:

$$\begin{aligned} M &= \{\text{model predicts site-present}\} \\ &\text{or} \\ M' &= \{\text{model predicts site-absent}\} \end{aligned}$$

The difference between these two sets of event classes is crucial for interpreting model results. Any single land parcel can be classified according to its condition in reality (S or S') and by its condition predicted by the model (M or M'). Because no model makes perfect prediction, the true condition and the model prediction of a land parcel may not agree. Comparing the relative values of S , S' , M , and M' provides a quantitative method for evaluating model performance. This notation will be referred to throughout the report.

3.1.3 Predictive Models as Decision Rules

An archaeological predictive model is a decision rule conditional on other, non-archaeological features of a location (Kvamme 1990:261). Decision rules can be generated using techniques ranging from an inductive analysis using statistical techniques to derive an equation from empirical patterns in sample data to a deductive analysis in which a trained archaeologist creates decision rules based on previous knowledge of cultural patterns. The critical question when constructing an archaeological model is the relative weights to associate with each non-archaeological variable, or in this case each environmental variable. A professional archaeologist working within a region will inevitably have a mental conception of where sites occur on the landscape. However, this information is often localized and may vary between archaeologists. The utility of statistical methods in model development relates to the independent method in which variable weights can be derived. Deductive knowledge is required for the initial variable selection, but the spe-

cific values for each variable are derived from the spatial patterns of the sample data. In this way a modeler can concentrate on macro-scale selection of variables and appropriate data structures and let the statistical method derive the micro-scale variable weights. Kvamme (1990), Carr (1985), and Parker (1985) provide a thorough review of various statistical and inductive methods. The predominant statistical technique used in archaeological predictive modeling is the logistic regression method. Logistic regression is discussed further in Section 3.5.

3.2 Factors Influencing Model Development

A survey of the available literature indicates predictive modeling has been utilized in various geographic and archaeological contexts (Lock and Stancic 1995; Westcott and Brandon 2000; Wheatley and Gillings 2002). The type of environment and cultural group under study influences model development, particularly in the selection of explanatory variables. Selection of relevant variables for model inclusion is dependent upon the mechanisms in which the cultural group under study interacted with the environment. Consider the differences between a nomadic hunter-gatherer on the Great Plains, a sedentary horticulturalist in the Blue Ridge Mountains, and a Roman agriculturalist in Italy. Clearly the relationship between cultural activity and the environment are different in these situations, thereby leading to a different set of relevant environmental variables selected for initial entry into the model. In terms of statistically evaluating model performance, the goodness-of-fit statistics designed to measure how well the archaeological data fit the input variables are affected by the number of variables selected for model inclusion. A condition known as *hyperfitting*, an upward bias of the goodness-of-fit statistic, occurs as more explanatory variables are added to the model (Kvamme 1988a). Therefore, the number of explanatory variables should be kept as low as logically possible.

A primary factor to consider in model design is the type and complexity of the economic system used by the cultural groups under study. Hunter-gatherer lifeways can be described as following an optimal food procurement strategy

in which the culture group extracts a living directly from the environment and patterns their site selection on the basis of minimizing energy output (Bamforth 1988; Butzer 1982; Jochim 1976). The close relationship between cultural behavior and environmental resources provides justification for the credibility of environmentally based model predictions. Market-driven economic systems of more advanced societies, primarily in Europe with some examples in North America, result in site patterns not entirely based on environmental resources. In these cases, social factors (distance to road, viewshed of defensive fortifications) may be important for describing site patterns (Kvamme 1990; Lock 2000; Lock and Stancic 1995; Wheatley and Gillings 2002). Spurious correlations may occur if inappropriate variables are included in the analysis. A model may be technically accurate but not have any real archaeological meaning; therefore, successful predictive modeling requires a theoretical understanding of the culture, environment, and time period under analysis.

Variation in environmental conditions, along with different social practices and economic systems between different cultural groups, requires adjustments to the modeling practice. Environmental considerations must be factored into the modeling methodology prior to analysis. One method of dealing with environmental variation is to divide the landscape into distinct physiographic regions and model each region separately (Hudak et al. 2000). Regional division of the landscape can be based upon any physiographic criteria, so long as the divisions represent significantly different resource zones. Distribution of resources within a region will influence site patterns within the region. If the distribution of resources or the type of resources change significantly between regions, then a model constructed for one region may not be appropriate for another region. It is important to note that if a quantitative method is used in the modeling process, the derived equation should only be implemented within the region for which it was developed.

3.3 Limitations

3.3.1 Garbage In – Garbage Out

Just as with other computer applications, the saying ‘garbage in – garbage out’ can be applied to predictive modeling. Incorrect input data, either environmental or archaeological, will adversely affect accuracy of the model output. Potential sources of error in archaeological data include the spatial position of site locations, the widespread lack of reliable cultural affiliation for site materials, and poorly distributed sample data. Potential sources of error in environmental data include issues of map accuracy and the inappropriate use of geographic data sets.

3.3.2 Environmental Determinism

Is it acceptable to predict human behavior using environmental variables? Have we not traveled down this theoretical road before? Geographers have spent the last century living down the philosophical implications of environmental determinism. Yet, when it comes to archaeological research, the temptation to use environmental models to describe cultural behavior is hard to resist. Many archaeological sites demonstrate repeated habitation, indicating the environmental resources of a location are found desirable by different cultures throughout time. If environmental resources are consistently found desirable, and those resources change slowly through time, then searching for unknown cultural materials on the basis of environmental conditions is justified. Well-constructed archaeological models accurately predict 70 - 85% of known archaeological sites. Repeated credibility of such accuracies indicates the relevance of predictive modeling as an investigative tool (Gaffney and van Leusen 1995; Hudak et al. 2000).

Opponents of predictive modeling point to the lack of theory behind the results of empirical models. One significant critique of the method relates to the central tenet of predictive modeling that environmental map data accurately represent the true landscape and that proximity measures derived from map data are important for explaining settlement strategies. Another critique of environmental data involves the reliability of modern maps to

accurately represent environmental conditions in the past, especially when considering existing paleo-climatic reconstructions (Ebert 2000; Gaffney and van Leusen 1995). While these critiques must be considered, predictive modeling as a tool for archaeological investigation should not be discarded. The changing nature of the landscape and accuracy of the available map data must be considered prior to model construction and when interpreting model output.

Others critiques point to the nature of the archaeological data and the fact that many archaeological sites are discovered through salvage efforts and, as a result, tend to be patterned. The manner in which sites are selected for inclusion in a predictive model must account for the sampling methods in which the original data were derived (Kvamme 1988b). Besides problems with how the spatial locations of sites are derived, precision in the ability to accurately assign a site to a specific cultural or temporal period is often lacking from archaeological databases. This problem is particularly acute in hunter-gather contexts in which cultural determination is often based on lithic technology. Use of stone tools for temporal classification requires diagnostic artifacts, which are often missing from site materials.

An important consideration of using predictive models is the amount of explanatory power given to the environmental correlations. Empirical correlations should be viewed as providing some insight into where cultural materials are located, not explicitly defining why the materials are there. The influence of human agency in cultural adaptation cannot be easily integrated into a numerical analysis of site patterns; however, archaeological discovery fundamentally requires new archaeological data for analysis and predictive models are effective tools for locating unknown cultural resources (Hudak et al. 2000; Warren and Asch 2000). Predictive models will not replace human investigation of the landscape; however application of a macro-scale landscape screening tool will improve research design, thereby resulting in more efficient archaeological surveys and cultural resource management (Verhagen 2000).

3.4 Use of Geographic Information Systems

Geographic Information Systems (GIS) are imperative to the construction, implementation, and testing of archaeological predictive models. Although archaeological predictive models have been created without the aid of GIS, the amount of map measurement required was prohibitive and resulting models typically had large land parcel sizes and could not easily be projected onto unknown areas (Pilgrim 1987). In the case of the current model, the study area occupies 4,134,650 land parcels. For each of these land parcels, six different variables were created, resulting in a total of 24,807,900 values for the environmental variables alone. This does not take into account the number of archaeological site locations, non-sites, and the number of calculations required to compute the statistical model. Obviously this amount of calculations could not be completed without the use of modern computers and GIS software. See Kvamme and Kohler (1988) for a thorough review of the use of GIS in archaeological predictive models.

3.5 Logistic Regression

The dominant method used in constructing quantitative archaeological predictive models utilizes a logistic regression technique, either binary or multivariate. Binary logistic regression, a type of probability model, is useful when the observed outcome is restricted to two values, which in this case represent the *site-present* {S} and *site-absent* {S'} event classes (Warren 1990). These events are coded as 1 and 0 respectively, for use in the database. Output of the binary logistic regression represents the probability of the event occurring, expressed as the *Prob(event)* or in this case the probability of a site occurring *Pr(M)*. In ordinary regression, the output value of the equation (Z) can be any value, positive or negative. Because the logistic model output is a probability, the output must be constrained between 0 and 1. Ordinary regression output (Z) must be converted to a probability value constrained between 0 and 1 (Clark and Hosking 1986). The standard linear regression equation can be generically described as:

$$Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_pX_p$$

where Z is the predicted output of the regression equation (dependent variable), B_0 is a constant term, B_p is a coefficient, and X_p is an independent variable for every variable in the equation. In order to convert the raw output to a probability of the event occurring, the following equation must be applied where e is the natural log and $(-Z)$ is the ordinary regression output multiplied by -1:

$$\text{Pr}(M) = 1 / 1 + e^{-Z}$$

And conversely, the probability of an event not occurring is expressed as:

$$\text{Pr}(M') = 1 - \text{Pr}(M)$$

Preference for logistic regression is based upon multiple factors. The method is robust with respect to the data normality and equality of variance assumptions required of related techniques, e.g., discriminant functions, and it can also handle nominal, ordinal, ratio, or interval level data (Gibbon 2000; Kvamme 1990; Parker 1985; Warren and Asch 2000). Kenneth Kvamme developed the method for use in archaeology in the early 1980's (Warren 1990); Kvamme's method of model development and assessment is used for the 2-D model described herein.

4.0 METHODOLOGY

Methods used for model development are divided into *database construction* (Section 4.1), *model development and results* (Section 5.0), and *model integration* (Section 6.0). Methodology used to create both models can be expanded to other areas of the Fort Hood Military Reservation; all the required GIS data layers either already exist or can be converted for GIS use within a short period of time.

4.1 Database Construction

The first step towards constructing an archaeological predictive model is the development of a GIS database suitable for predictive modeling. In this case, the model database contains geographic information about the location of archaeological sites, terrain, hydrology, and geomorphology. To be included as an independent variable in the predictive model,

the particular landscape characteristic must be relevant to explaining cultural behavior and structured for use in the GIS.

The majority of the geographic data used in this model are derivatives of geographic datasets provided by the Center for Environmental Management of Military Lands (CEMML) or Fort Hood. GIS data specifications are explicitly described in Appendix A: GIS Data Layers.

4.2 Archaeological Database

Unique features of the Fort Hood archaeological database include the large number of recorded sites, the classification of those sites based on eligibility for the National Register of Historic Places (NRHP), and the 100% site survey of the study area. Each of the features has implications for the specific methods used to construct the 2-D model and will be discussed throughout the report.

Structurally, the archaeological site database consists of a polygon shapefile of known archaeological site boundaries and a collection of randomly generated point locations that represent known non-sites. Data layers were converted from vector to raster format for use in the model.

4.2.1 Site Data

Construction and testing of a 2-D statistical model require a sample of archaeological sites for training the model and a separate sample for testing the model. The sample of sites used to construct the model requires particular attention; consideration must be paid to the composition and distribution of the sample. Archaeological characteristics of the sample determine the type of site predicted in the model output. For example, if the sample is composed of surface finds from the prehistoric period, model output will emphasize surface locations with environmental conditions similar to those of the prehistoric sites.

For this model, the total number of sites was split into two separate shapefiles on the basis of eligibility for the NRHP. A total of 118 sites were divided into a NRHP eligible training sample (44 sites) and a NRHP non-eligible testing sample (74 sites). Once rasterized, the training sample contained 56,206 land parcels and the testing sample contained 82,275 land

parcels. Figure 5 displays the distribution of the training site samples. The known site testing sample is discussed in the Model Assessment section.

4.2.2 Non-Site Data

Binary logistic regression compares the environmental patterns of known archaeological sites {S} with the patterns of known non-sites {S'}; therefore, a set of known non-sites must be included in the model training sample. The important consideration in constructing the non-site sample is to adequately capture the range of landscape variability. In order for the model to differentiate the landscape effectively, training data must reflect all the potential landscape decisions that could have been chosen for cultural activity. If the non-site sample does not contain the range of landscape variation, the model output will not reflect the true landscape preferences exhibited in the spatial pattern of sites.

Non-site data used in the model consisted of 435 points randomly distributed throughout the study area. Points were buffered to a radius of 40 meters and rasterized for use in the model (Figure 6). Justification for the random point method is based on the claim of the Fort Hood Cultural Resources Office of a 100% site survey. Consequently, if a site does not exist at a specific location in the database, the assumption is that in reality no site exists at that location. Non-site points were located within the study area boundary and outside any existing site boundaries. The appropriate ratio of sites to non-sites is not agreed upon in the literature, and ratios ranging from approximately 1:2 to 1:10 are reported (Kvamme 1992; Warren and Asch 2000). Warren (1990) provides a cursory discussion of the implications of enlarging the non-site sample relative to the site-sample. The final number of non-site parcels was designed to reflect a 1:1.5 ratio between sites and non-sites. A second set of randomly generated non-site points was used for model assessment. Details concerning the construction of the non-site testing sample are described in the Model Assessment section.

4.3 Environmental Database

Determining the appropriate variables to use in the predictive model is the first step in constructing the environmental database. In their native format, landscape metrics contained in standard GIS datasets are of limited value to predictive modeling. Most environmental variables used in predictive modeling are derivatives of standard GIS datasets. Consider a Digital Elevation Model (DEM) as an example: the DEM is an extremely useful tool for visualizing landscape variation across space, however, simple elevation values are not a powerful variable for predicting archaeological site locations. Although elevation values themselves are not powerful variables, the DEM can be used in combination with various

computer algorithms to derive landscape variables significant for predicting human behavior such as slope, shelter, and relief.

4.3.1 High Resolution Terrain Dataset

Baseline terrain data provided for the project is a DEM with a 5m² pixel resolution (Figure 3). Data resolution of this scale is only possible using a high resolution sampling method; in this case the DEM was created by interpolating a vector contour map with a 3-meter contour interval derived from aerial photogrammetry. Specifics for this dataset can be found with the metadata provided by CEMML. The accuracy and small pixel size of the DEM model provides a unique opportunity for the creation of other landscape variables to include in the

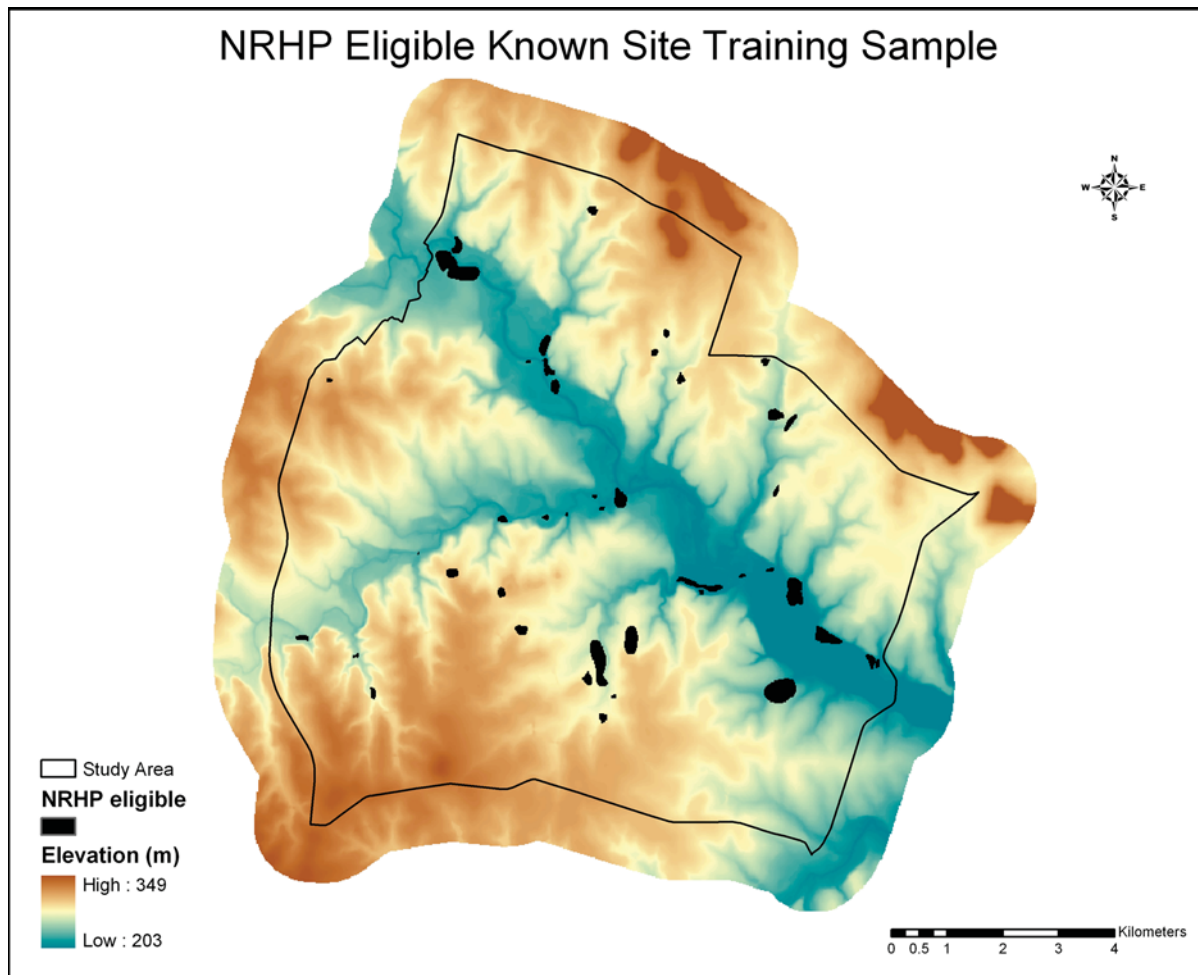


Figure 5. NRHP Eligible Known Site Training Sample

model. Current literature indicates most predictive modeling projects use the USGS 30m² DEM. Any derivative layers created from the 30m² DEM subsequently have the same 30m² grid cell size. The very high resolution of the Fort Hood DEM means that any derivative layers created from it also have the 5m² pixel size. Therefore, derived slope, shelter index, and relief variables are very high resolution. The terrain variables included in the 2-D model include *slope*, *relief within a 150-meter radius*, *relief within a 300-meter radius*, and a '*shelter index*' within a radius of 150 meters. Figures for the environmental variables are found in Appendix A: GIS Data Layers.

4.3.2 Hydrographic Dataset

Another example of the need to transform a typical dataset for use in a predictive model involves hydrographic datasets. Typically, these datasets include line representations of rivers and streams, usually coded with a form of stream classification. The fundamental question posed by the model requires a different type of data than that provided by the typical hydrographic dataset in its native format. A standard hydrographic dataset indicates where a stream is located, but the model requires information about how far each land parcel is away from a stream. Therefore, the original hydrographic dataset must be used to derive a secondary dataset in which every land parcel

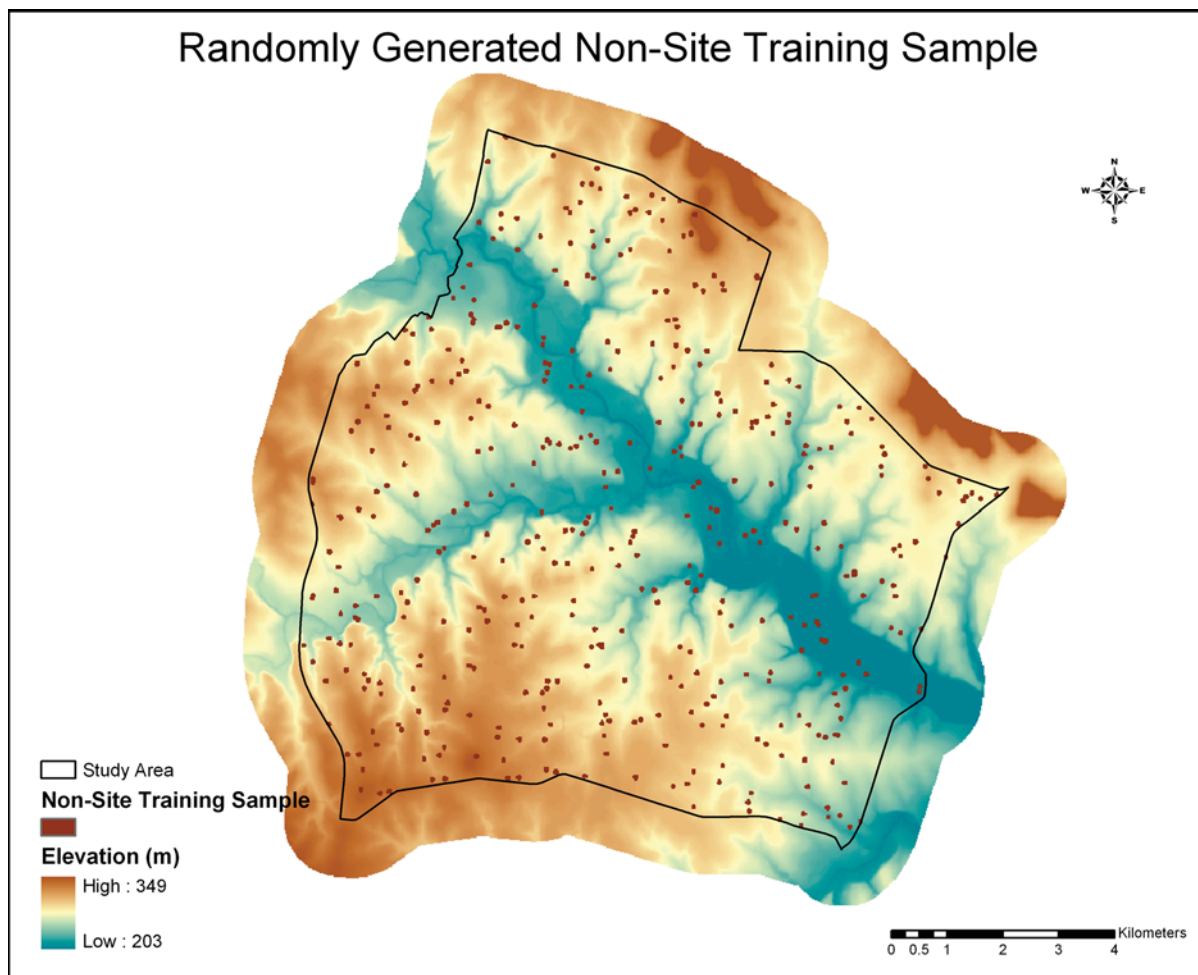


Figure 6. Randomly Generated Non-Site Training Sample.

contains a value describing the distance to the nearest water source.

Hydrologic data used in this model, provided by CEMML, consisted of a line shapefile representing the location of streams on Fort Hood. No attribute data were associated with the files. Stream segments were manually separated into primary and secondary streams (Figure 7). Using the Spatial Analyst extension, two “Distance from...” grids were created in which each 5m² grid cell contains a value for the distance to the closest primary and secondary stream respectively. The two “Distance from...” grids comprise the hydrologic component of the model. Figures depicting the “Distance from...” grids are located in Appendix A: GIS Data Layers.

Springs are a fairly ubiquitous feature of the Fort Hood landscape and most likely had an influence on cultural behavior. However, no comprehensive database of spring locations is available. It is difficult to attempt the inclusion of an environmental variable in a predictive model if the dataset is incomplete. Previous modeling experience with springs data has shown that including a non-comprehensive spring dataset will skew model results.

4.3.3 Geomorphic Dataset

No formal geomorphic variable was used in the construction of the 2-D model, but digital geomorphic data were created for use in the 3-D model. The primary sources of geomorphic data were revised digital versions of alluvial land-

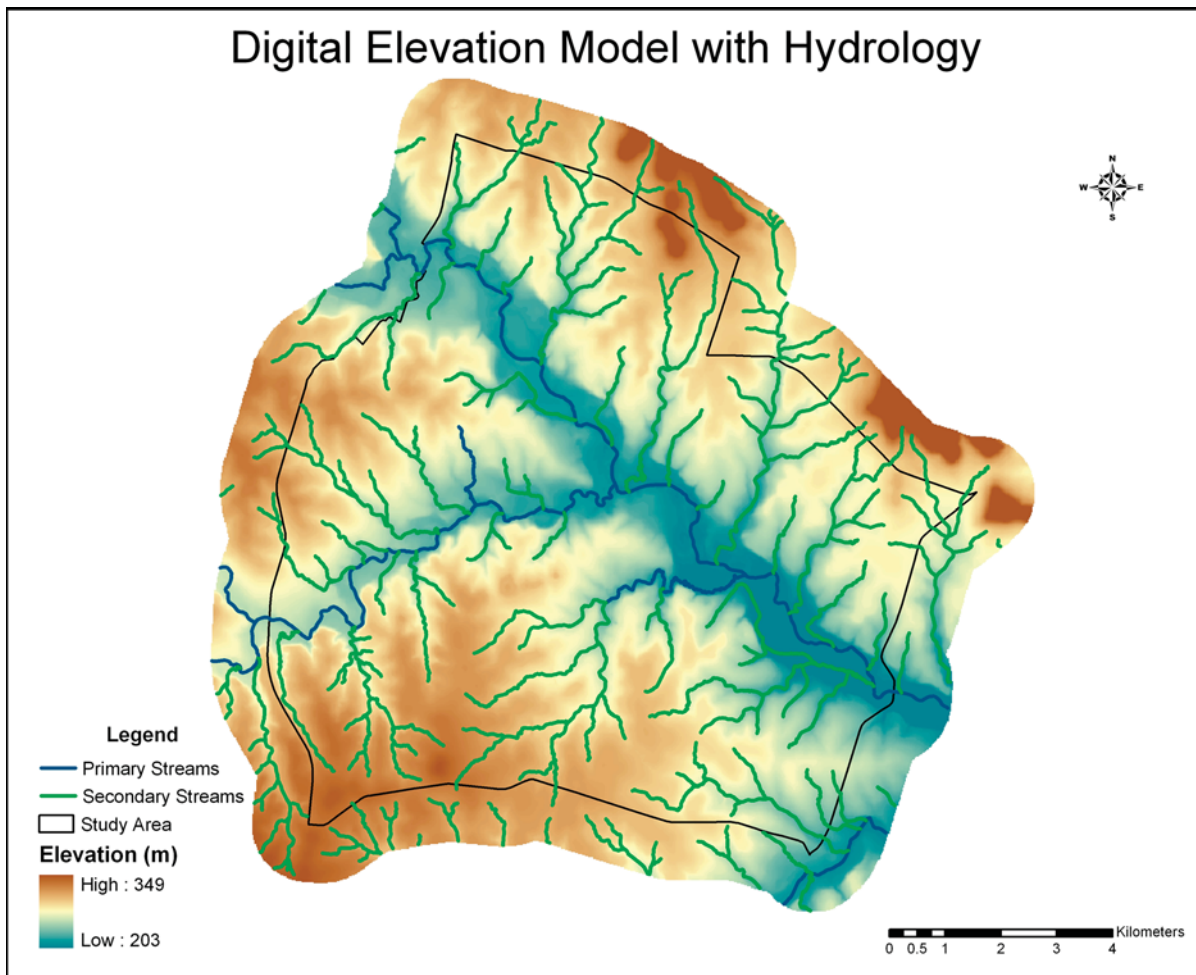


Figure 7. Digital Elevation Model with Hydrology.

form maps produced by Nordt (1992). Although CEMML provided a geomorphic map based upon a SSURGO reclassification, coarse resolution of the data rendered it ineffective and digital versions of the published analog maps were created. Using Nordt's maps as a guide, the T0-T1-T2 terrace landforms of Cowhouse Creek and Table Rock Creek were re-mapped, in a GIS environment, using a combination of data layers including a high-resolution DEM, derived slope data, high-resolution aerial imagery, and existing 1:24,000 topographic maps (Figure 8). The resulting dataset forms the basis of the 3-D model.

5.0 MODEL DEVELOPMENT AND RESULTS

The two models discussed herein represent two distinct approaches to predictive modeling, use different input data, and produce different results. Taken together, these models provide a unique approach to predicting the location of unknown cultural material. The basis of the 2-D predictive model is to quantitatively relate the presence or absence of archaeological material to the environmental characteristics present at a location. This model deals only with the surface and makes no consideration of the potential for deeply buried sites. The 3-D model deals with the stratigraphic composition of alluvial landforms and the implications for preservation of cultural material on the landform

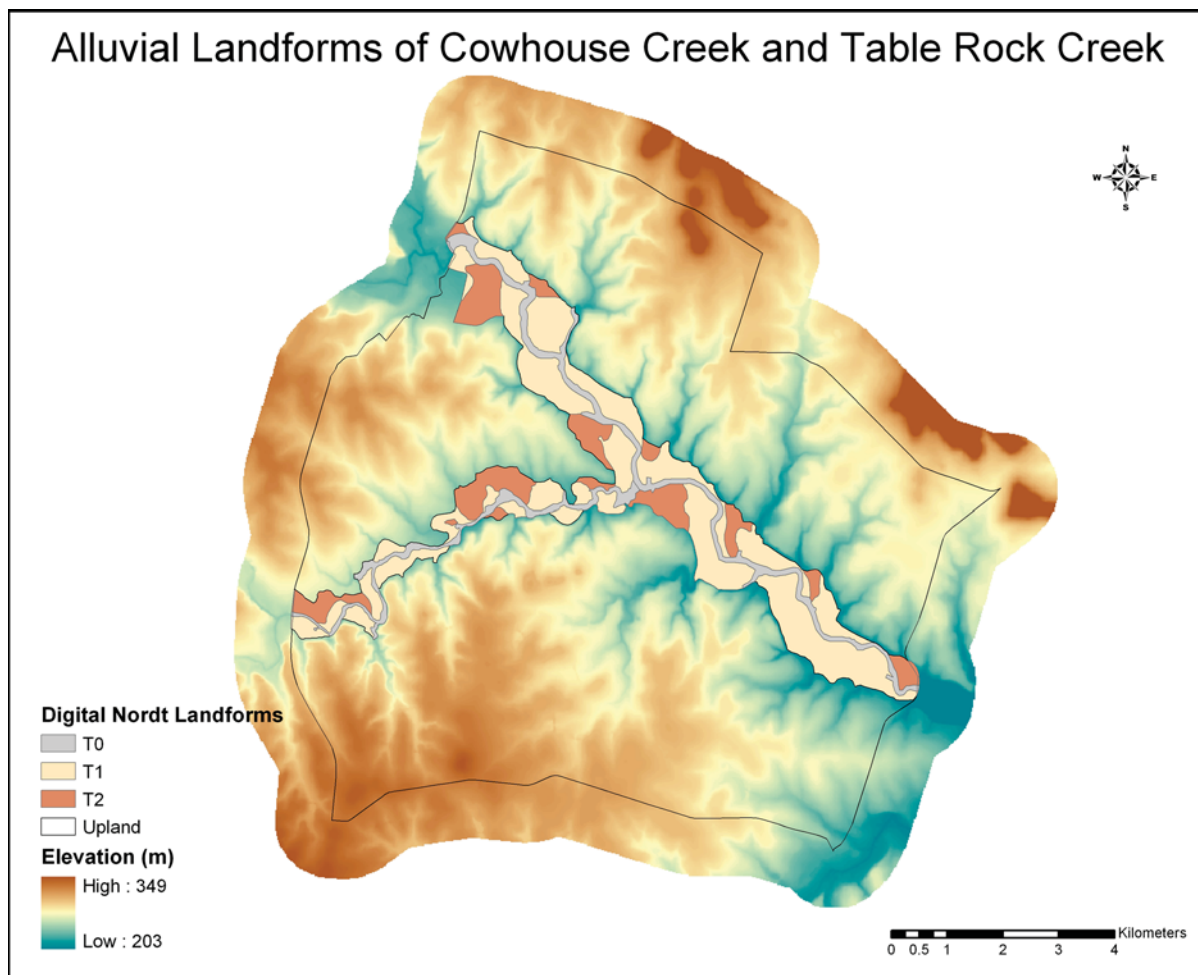


Figure 8. Alluvial Landforms of Cowhouse Creek and Table Rock Creek

surface and at depth. The 3-D model requires stratigraphic landform data and is currently limited to the alluvial areas of Cowhouse Creek and Table Rock Creek.

5.1 Statistical (2-D) Model

The basic goal of a predictive model is to classify a given land parcel into a site-present or site-absent class based upon measurable landscape characteristics, e.g., distance to water, local relief, and slope. The statistical method used to determine the relationship between cultural material and the geographical characteristics of the location is the binary logistic regression analysis. Landscape data for known site areas and known non-site areas are extracted from the GIS and entered into SPSS statistical software for analysis. Once an equation is developed in SPSS, it is re-entered into the GIS and “mapped” across the landscape, meaning that all 5m² grid cells receive a probability value output from the logistic regression equation. The resulting model is a GIS raster data layer that represents a continuous probability surface for encountering cultural material. The model is developed using a set of land parcels known to contain cultural material and a set of land parcels that do not contain cultural material. The developed model is tested against a set of known archaeological locations that were withheld from the original model development. Relative stability of these landscape features justifies the use of modern map data within the model. Ability of the model to predict the ‘testing’ sample of sites determines the power or accuracy of the model.

Variables used in the model include *slope*, *relief within a 150-meter radius*, *relief within a 300-meter radius*, and a *‘shelter index’ within a radius of 150 meters*. Selection of environmental variables was designed to reflect components of the landscape significant for a hunter-gather subsistence strategy (Kvamme 1992). These variables are derivatives of modern map data that were created and stored within a GIS. Although landscapes change over long periods of time, this set of environmental variables was selected because they represent reasonably stable features during the last 15,000 years. Areas of the landscape that undergo relatively rapid geomorphic change, specifically the alluvial systems, are dealt with in the 3-D model.

Training the predictive model requires input data for known site land parcels. The known site parcels were selected from the set of sites eligible for the National Register of Historic Places (NRHP). These site locations are eligible for the NRHP on the basis of site preservation and diagnostic artifact collections. Locations of the NRHP-eligible sites may reflect a geologic tendency for preservation. By using these sites the model incorporates landscape positions favorable for cultural activity and site preservation. Model output could therefore be viewed as predicting areas of the landscape that have a high probability of containing *preserved* cultural material.

Binary logistic regression also requires data about the null condition, which in this case is known non-site locations. Selection of the non-site sample must reflect the true variability of the landscape. The model must contain information about all possible landscape locations. A true representation of the landscape in the non-site data allows the model to better discriminate locations of cultural activity from the overall set of landscape choices.

Statistical information required to construct the predictive model was extracted from the GIS and imported into an SPSS database which formed the basis of the statistical analysis. In total, 56,206 known site parcels and 86,237 known non-site parcels were extracted from the GIS for the training sample. An additional set of 82,275 site parcels and 8,407 non-site parcels were extracted for the testing sample.

Due to the large size of the training samples, the statistical calculations used in model construction were based upon a 10% sample of the training data exported from the GIS. The small land parcel size resulted in redundant statistical data collection and the number of samples was too large to compute the logistic regression model. Although only 10% of the cells extracted from the dataset were used for model construction, visual analysis of the spatial pattern of the 10% indicate they were randomly extracted from the overall set and therefore represent a valid statistical sample. The number of samples in the 10% extraction is still significantly large (8,593 non-site and 5,532 site cells).

A backward conditional stepwise logistic regression technique was used to create the

predictive model. Using this method, all the variables are initially entered into the regression, and insignificant variables are removed at each iteration of the model. In this case, all variables were found to be significant, and none were removed from the model. In order to test if the backward conditional regression was valid, a model was also constructed using a forward conditional stepwise logistic regression. This method differs from the backward conditional in that no variables are included at the first step of the model; additional variables are added at each iteration based upon the explanatory power of the variable. In the forward conditional model all variables were found significant and the variable coefficients were identical to the backward conditional method. No coefficient was included in the regression equation.

5.1.1 Model Output

The regression equation developed by SPSS is expressed as the following:

$$Z = (-.0256979 * \text{SLOPE}) + (.0000084 * \text{SHELTER150}) + (.0367382 * \text{RELIEF150}) + (.0206011 * \text{RELIEF300}) + (-.0010430 * \text{D_MAIN}) + (-.0033151 * \text{D_SECONDARY})$$

In order to convert the Z equation into a probability score, the following equation is also required:

$$\text{FinalModel} = 1 / (1 + \text{EXP}(-Z))$$

Using the model coefficients derived in SPSS, both equations were re-entered into the GIS. Using the Raster Calculator function within Spatial Analyst, the equation representing the model output was entered. The GIS then applies the equation to every land parcel in the study area. The landscape characteristics of each land parcel are used in the equation and a probability score for each land parcel is generated. The resulting product is a decision surface (FinalModel) describing the probability for the land parcel to contain cultural material (Kvamme 1992).

Model output for each land parcel is a value ranging between 0 - 1. Values near 1 are associated with land parcels with environmental conditions similar to the site-present event

class. Low values, near 0, have environmental conditions similar to the site-absent event class. Values in the middle are 'indeterminate' (Kvamme 1992:30; Warren 1990). See Figure 9 for a visual depiction of model output. For display purposes, the raw probability values were converted to whole numbers with values ranging between 0-10 and classified into 10 equal classes.

5.1.2 Model Assessment

Model accuracy is assessed using the methods describe in Kvamme (1992). Methods and logic for the accuracy assessment are reported below. The optimal modeling goal is to maximize the percentage of correctly classified site-present class (S) in a minimum of land area (M). The techniques for calibrating the model for this goal are a critical component of model assessment.

Accuracy of the predictive model is measured primarily in terms of its ability to correctly classify both known site locations and known non-sites. A complete representation of model accuracy includes both the percentage of correctly identified sites and percentage of correctly identified non-sites. The percentage correct of sites represents the percentage of sites (S) that are correctly classified within the site-present class of the model (M), and the percentage correct of non-sites (S') represents the percentage of the site-absent class correctly classified in the site-absent class of the model (M'). These two measures can be described as $100\text{Pr}(M|S)$ and $100\text{Pr}(M'|S')$. Additional assessment metrics include the probability of a site occurring when the model predicts a site, $\text{Pr}(S|M)$, and the probability of a site occurring when the model does not predict a site, $\text{Pr}(S|M')$.

Kvamme (1988a) indicates a predictive model must perform better than a random chance model. Using the metrics described above, and the base-rate probabilities, the model can be evaluated in a quantitative and defensible manner. Comparing the measures of model accuracy with the base-rate probabilities provides a method of quantifying model accuracy as a percentage increase over random. Computation of the random chance or base-rate models are discussed below.

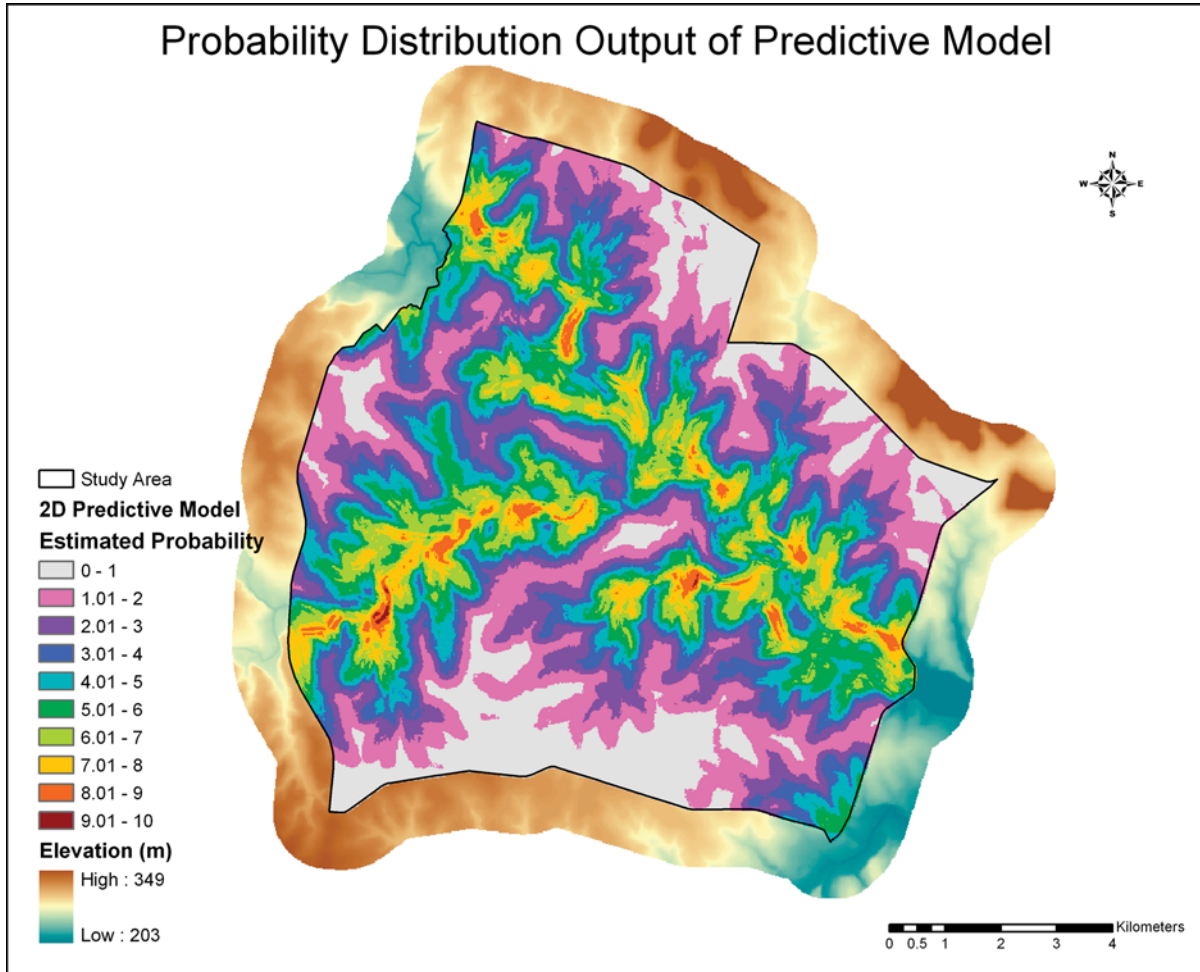


Figure 9. Probability Distribution Output of Predictive Model.

Output of the model classifies the landscape into two event classes (M and M'), yet the output of the regression is a probability score ranging from 0 - 1. A 'cut-point' in the range of probabilities must be established. For example, the standard cut-point is 0.5, meaning that any land parcel with a probability score of 0.5 or greater would be assigned to the site-present (M) class and any score less than 0.5 would be in the site-absent (M') class. This relationship is described mathematically as:

$$\begin{aligned} M &= L \geq 0.5 \\ &\text{and} \\ M' &= L \leq 0.5 \end{aligned}$$

where L is the decision point or cut-point at which the range of values is divided. Although

0.5 is the standard cut-point, the value can be shifted higher or lower based on modeling needs. Consider if the cut-point were moved or 'slid down' to 0.4, the percentage of archaeological locations correctly identified would increase, but an associated decrease occurs in the percentage of non-site locations correctly identified. The change in percentage correctly identified is due to a larger land area being included in the site-present class (M) as the cut-point is lowered. Using this logic, it is possible to correctly identify 100% of the archaeological sites by moving the cut-point to an extremely low number (0.01); however, the model would accurately predict 0% of the non-site locations and the site-present class (M) would occupy 100% of the landscape. This would offer no utility to land use managers.

Two methods for determining the appropriate cut-point are found in the literature. The first places the cut-point at the graphical intersection of the site-present and site-absent classes. The intersection cut-point represents the model optimum, the cut-point in which the greatest percentage of site-present and site-absent parcels are correctly classified simultaneously (Warren 1990). Kvamme (1992) indicates that a predictive model should correctly identify at least 85% of the site-present sample. Therefore, the cut-point is established by determining the value at which 85% of the sites are correctly classified. The Mn/Model goes a step further in requiring that their Phase 3 models correctly identify 85% of the sites and that the landscape area classified as site-present (M) does not occupy more than 33% of the total landscape (Hudak et al. 2000). For the purposes of this model, the cut-point is set at the level in which approximately 85% of the sites are accurately classified.

5.1.2.1 Base-Rate Probabilities

A fundamental requirement of quantitative model assessment is computation of the base-rate, or random chance, probabilities. A total of 118 sites are located within the study area, 44 NRHP eligible and 74 NRHP non-eligible sites. These sites occupy a total of 138,481 5m x 5m land parcels. The entire study area occupies 4,134,650 land parcels. The base-rate or a priori probability of the site-present {S} event class can be calculated as:

$$\Pr(S) = 138,481 / 4,134,650 = 0.0335$$

and the site-absent class {S'} as:

$$\Pr(S') = 3,996,169 / 4,134,650 = 0.9665$$

The event classes are mutually exclusive and represent all possible outcomes, i.e., $\Pr(S) + \Pr(S') = 1$. The base-rate probabilities provide "pure-chance" probabilities for each archaeological event class. Using an example from Kvamme (1992), the "pure-chance" probabilities are analogous to the probability of identifying a site by throwing darts at a map. By chance, 3% of the darts would land on a site parcel and 97% would not. Establishing the a priori probabilities for the two event classes sets the stan-

dard by which the predictive model is evaluated. In order to be considered effective, the model must "predict an event occurrence with probability greater than the event's base-rate chance of occurrence" (Kvamme 1992:28). Written mathematically, the previous statement is expressed as:

$$\Pr(S | M) > \Pr(S)$$

where $\Pr(S | M)$ is the probability of a site given that the model specifies a site. The mathematical expression is the quantitative version of the statement that a model must perform better than random chance.

Calculation of $\Pr(S | M)$ was designed to be conservative due to the inclusion of all known site-present parcels in calculation of the base-rate probability, $\Pr(S)$. If the base-rate probability was computed using only the training sample, $\Pr(S)$ would equal 0.0135, significantly lower than the value of $\Pr(S)$ equal 0.0335 used in assessment calculations. Lowering the value of $\Pr(S)$, and the subsequent increase of $\Pr(S')$ would not make the model more powerful in reality, but the statistics used to calculate the model's predictive power increase over random chance would increase.

5.1.2.2 Training Data

The model was trained with a 10% sample of the NRHP-eligible site locations and randomly generated non-site locations. The sub-sampling was required due to the large number of overall samples; SPSS would not compute the regression equation using all the land parcels. However, preliminary model assessment utilized all the available site and non-site training parcels. Histograms displaying the distribution of both site and non-site training parcels along the probability gradient are shown in Figures 10 and 11, respectively. Optimally, the site sample will cluster near the high end of the probability range and the non-site sample will cluster near the low end of the range. Clustering of the two event classes at different ends of the probability spectrum indicates the model is effectively separating the landscape.

Histograms indicate that both samples are effectively separated along the probability spectrum. One potential problem relates to the large number of known site sample cells in the

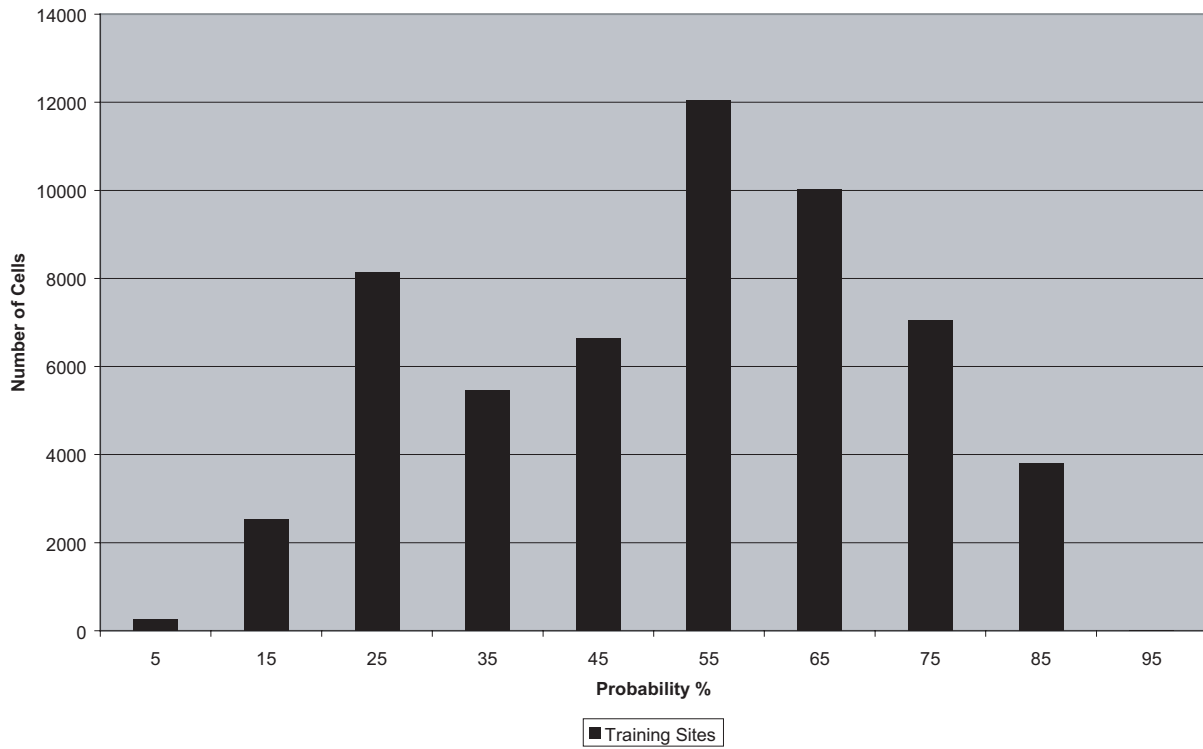


Figure 10. Predictive Probability of NRHP-Eligible Site Training Sample.

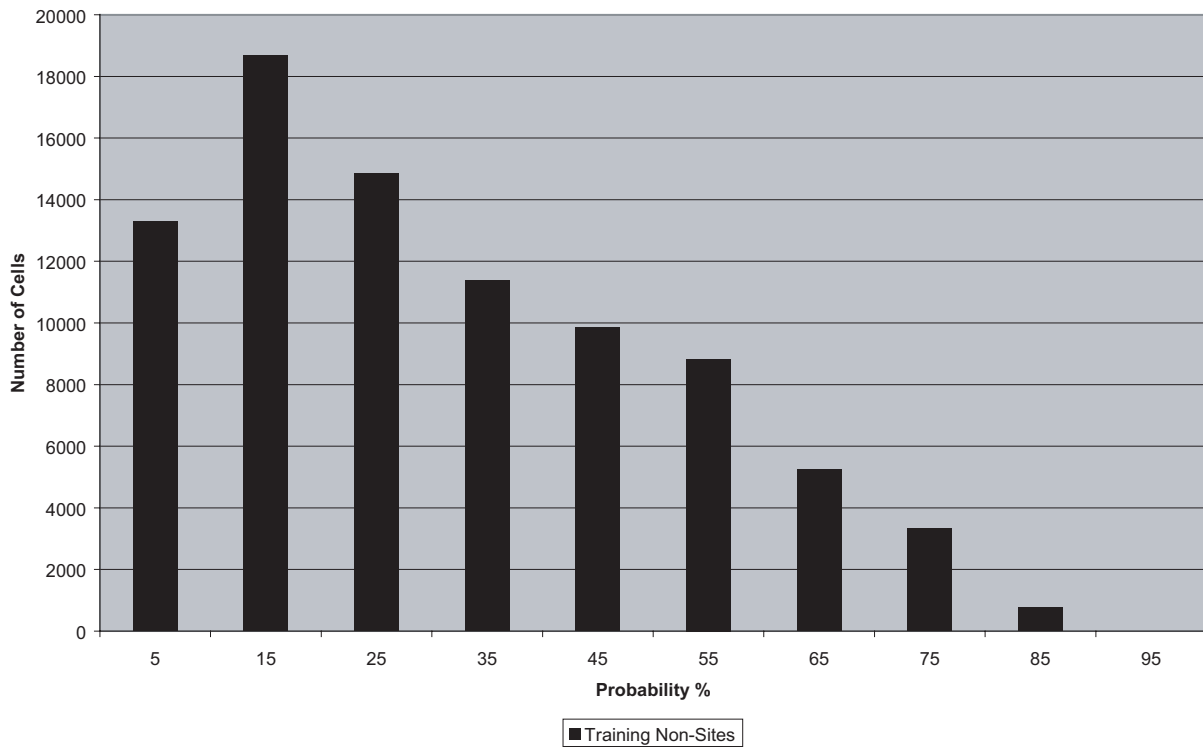


Figure 11. Predictive Probability of Known Non-Sites Training Sample.

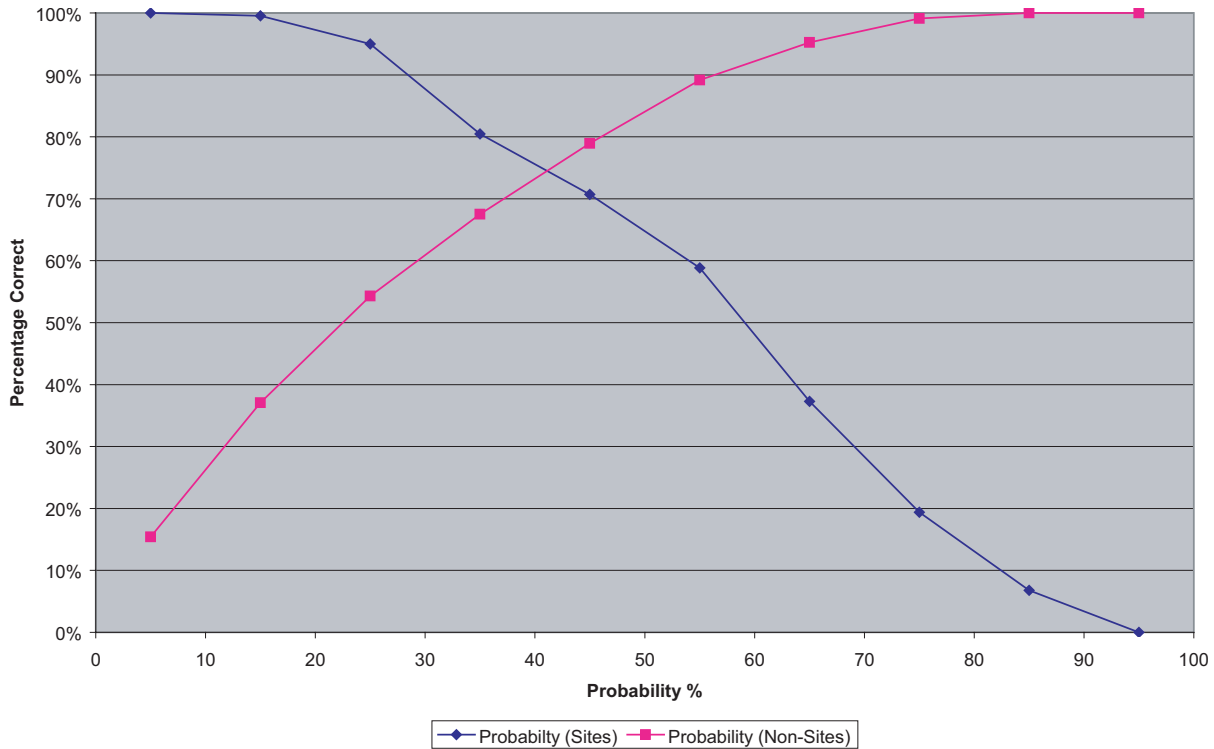


Figure 12. Optimal Cut-Point of Predictive Model (training data).

21-30% probability class. In order to correctly classify 85% of the known sites, the cut-point must be shifted significantly to lower than the standard 0.5 value. In this case, the cut-point is set a 0.25. Figure 12 displays the results of the graphical cut-point. The 0.25 cut-point is well below the ‘optimal’ cut-point, located at the graphical intersection of the training and testing samples. In the case of the training data, the optimal cut-point is approximately 0.42. However, using this value will only correctly classify 75% of the known sites.

Table 1 displays the accuracy assessment using a cut-point = 0.25.

Assessment of the training data at the 0.25 cut-point indicates the model correctly classifies over 86% of the sites and 54% of the non-sites. Alternatively, 13% of sites and 46% of non-sites were incorrectly classified. The reported accuracy of the site-present class (M) most likely contains some upward bias due to the use of training data, therefore an additional set of data were used to more rigorously analyze model output (Kvamme 1992).

Table 1. Accuracy assessment using a 0.25 cut point

True Condition	Model Predictions		Training Data
	M (Site-Present) (L .25)	M' (Site-Absent) (L .25)	
S (Site-Present)	48779 (86.79%)	7427 (13.21%)	56206
S' (Site-Absent)	39677 (46.01%)	46560 (53.99%)	86237

5.1.2.3 Testing Data

In order to better assess model accuracy, an additional set of known sites were withheld from model development and tested against the completed model. The set of land parcels from the NRHP non-eligible sites were used for secondary model testing. A total of 82,275 known site parcels were extracted from the NRHP non-eligible sites for use as a testing sample (Figure 13). A separate random sample of non-sites was also created. Identical procedures for creating the random sample were used, with the exception that the points were buffered to a distance of 15 meters. A smaller buffer size results in 8,407 non-site land parcels when the points were converted from a vector to raster data format (Figure 14). The smaller number

of parcels was found to be acceptable due to the high degree of similarity with the larger training non-site sample. Similarity between the two samples indicates the utility of the random point method of generating a sample of non-sites for use in the model.

Histogram distributions of the site and non-site testing sample are shown in Figures 15 and 16 respectively. Probability distributions are similar to the training samples, including the disproportionately large number of site parcels in the 21-30% probability class. As in the training sample, the large number of known site land parcels in a low probability range requires that the cut-point of the model be moved lower on the probability scale. The cut-point must again be set at 0.25 to correctly classify 85% of the site testing sample. The 0.25

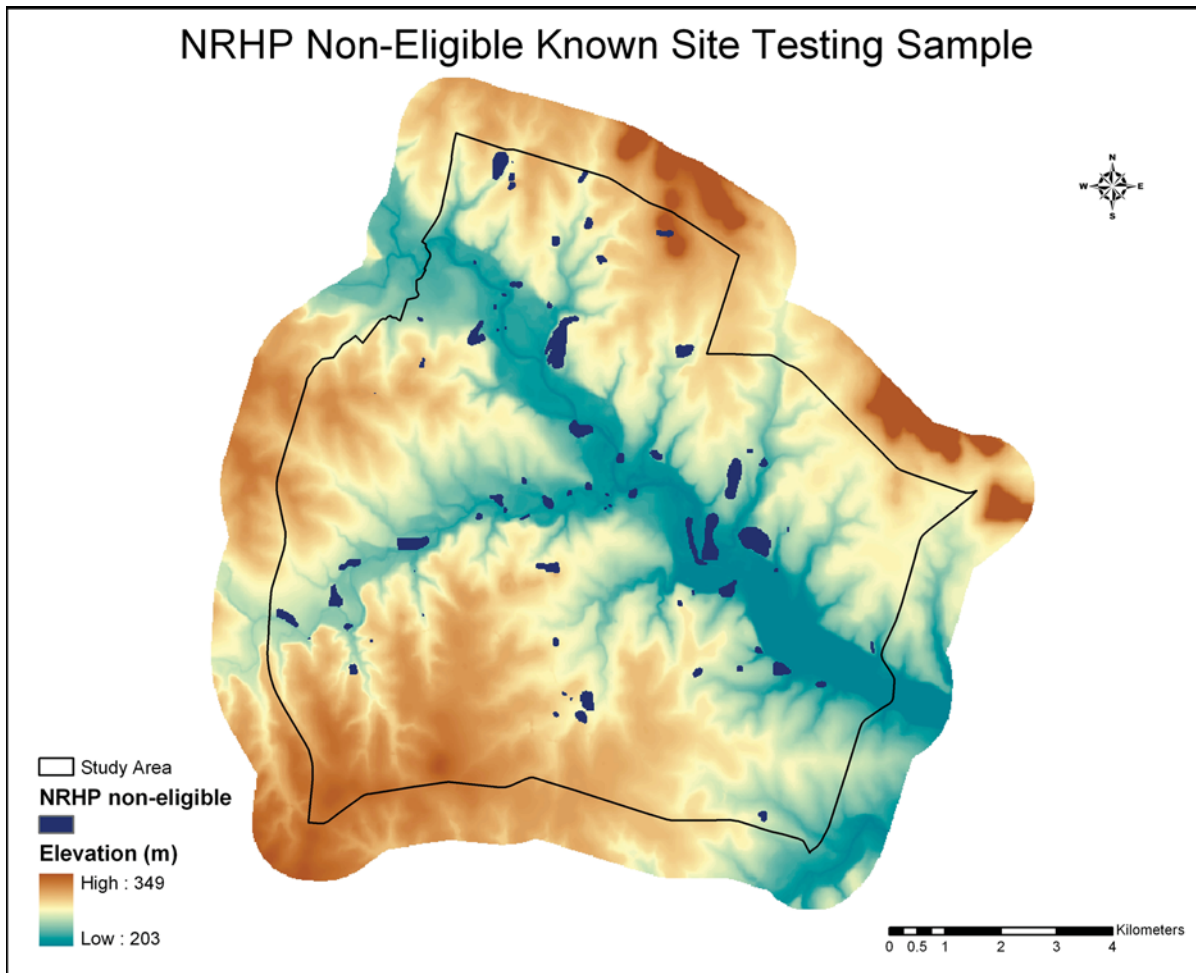


Figure 13. NRHP Non-Eligible Known Site Testing Sample.

cut-point value is again well below the ‘optimal’ cut-point derived from the graphical intersection method (Figure 17).

Table 2 displays the results of the testing sample at the 0.25 cut-point level.

Predictive power of the model at the 0.25 cut-point diminishes slightly with the use of testing data. This drop could be due to the nature of the NRHP non-eligible site locations or to the accuracy inflation associated with the use of training data to assess model accuracy. In either case, 83% of the testing sample site data and 51% of the non-site testing data are correctly classified. Additional assessment metrics can be derived by comparing the total amount of land parcels with the amount of known site parcels classified within a particular probability class (Figure 18). If the curves of the known

site samples have higher values than the overall landscape, the model is performing better than random chance. By manipulating the data used in the construction of Figure 18, it is possible to exactly quantify the model’s percentage increase over random for any given probability class.

Graphical analysis indicates that by classifying 85% of the known sites into 56% of the land area, the 0.25 cut-point value represents approximately a 30% gain over a random classification (Figure 19). Dividing the study area into the site-present class (M) and the site-absent class (M') at the 0.25 cut-point results in 56% of the study area assigned to the site-present class (M) and 44% assigned to the site-absent class (M') (Figure 20).

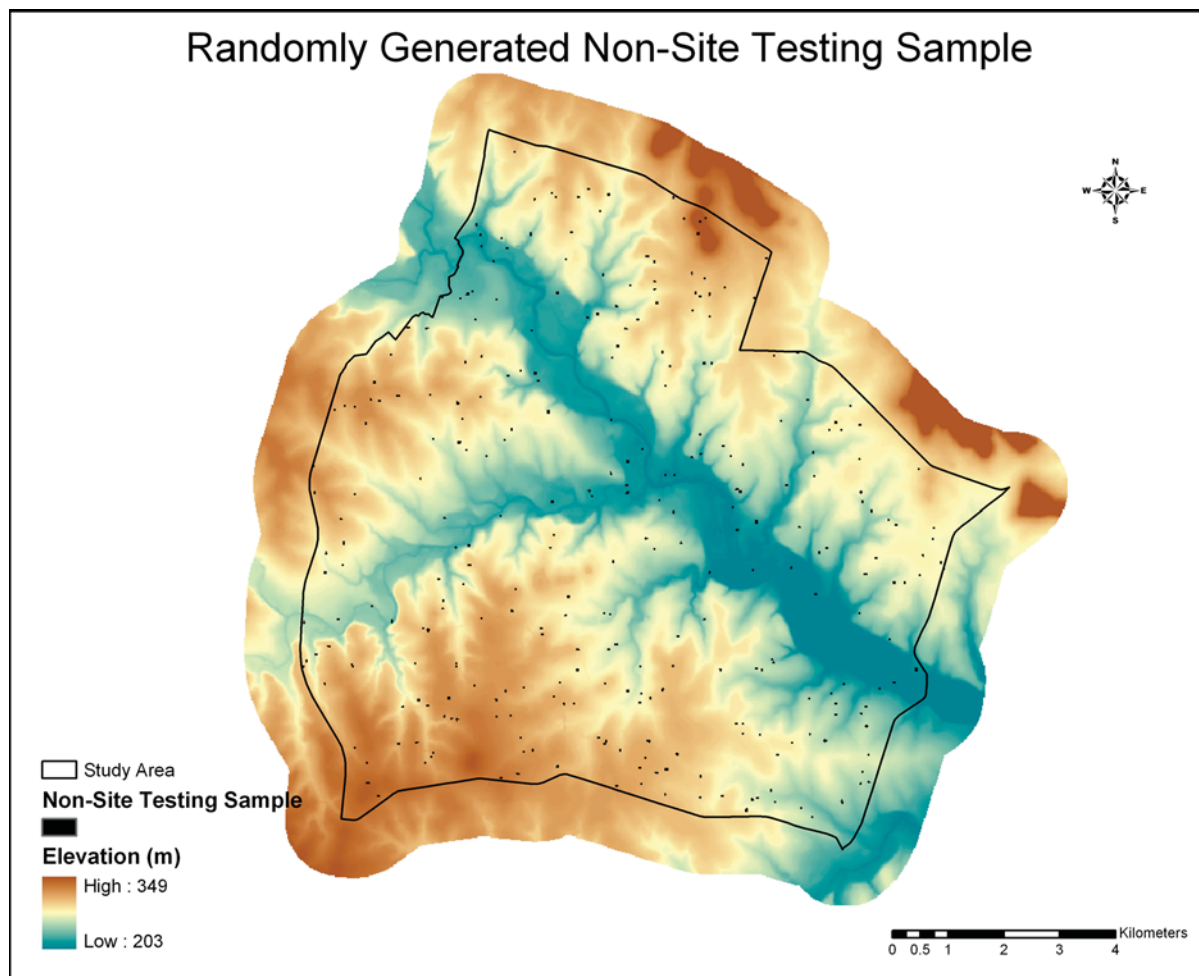


Figure 14. Randomly Generated Non-Site Testing Sample.

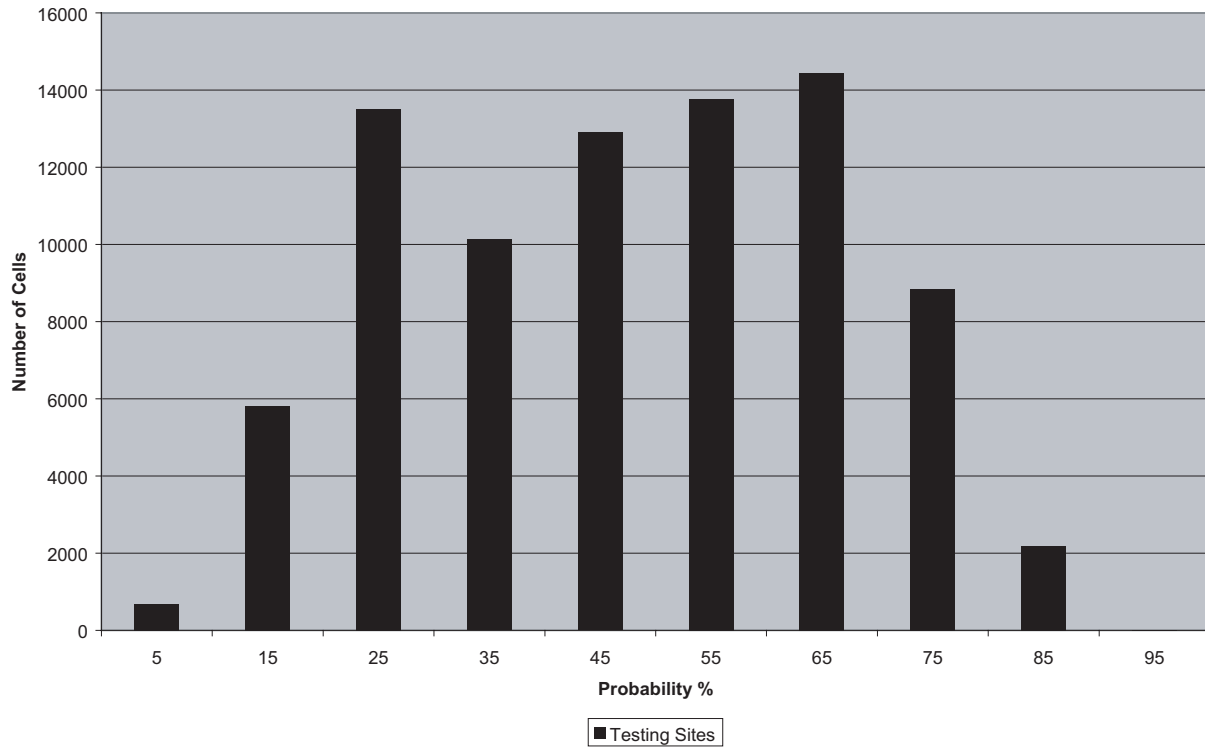


Figure 15. Predicted Probability of NRHP Non-Eligible Testing Sample.

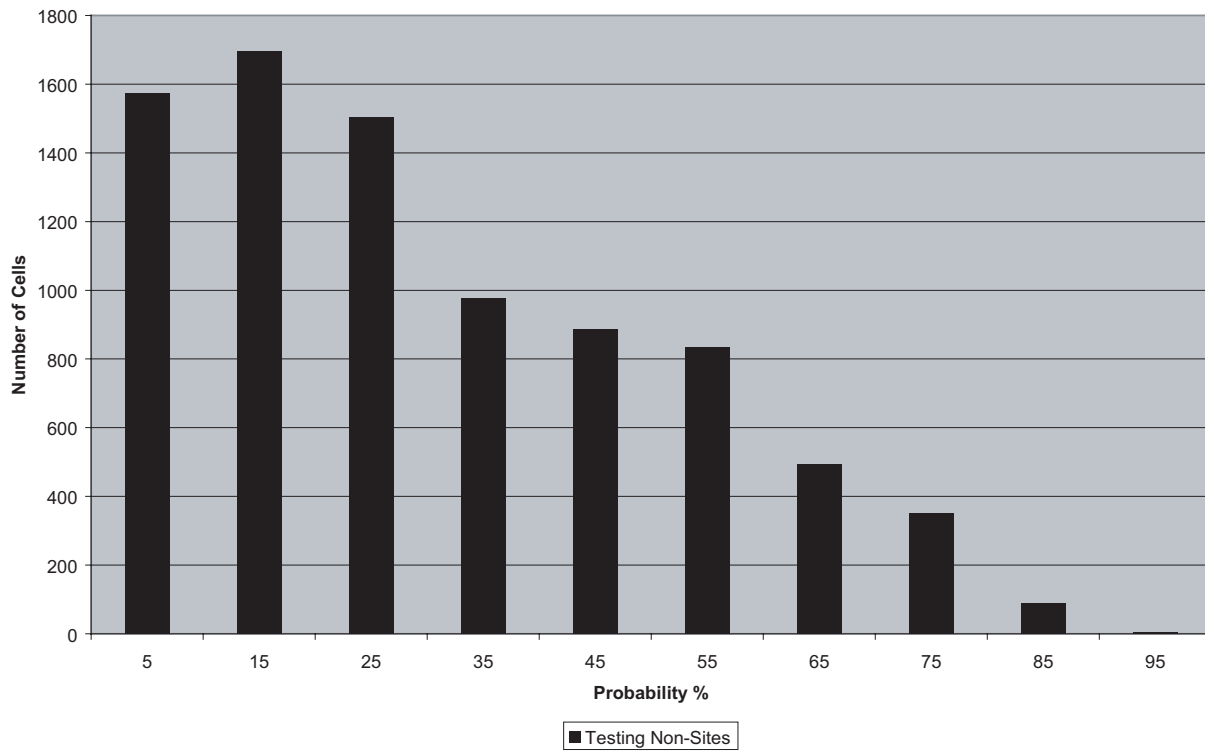


Figure 16. Predicted Probability of Known Non-Site Testing Sample.

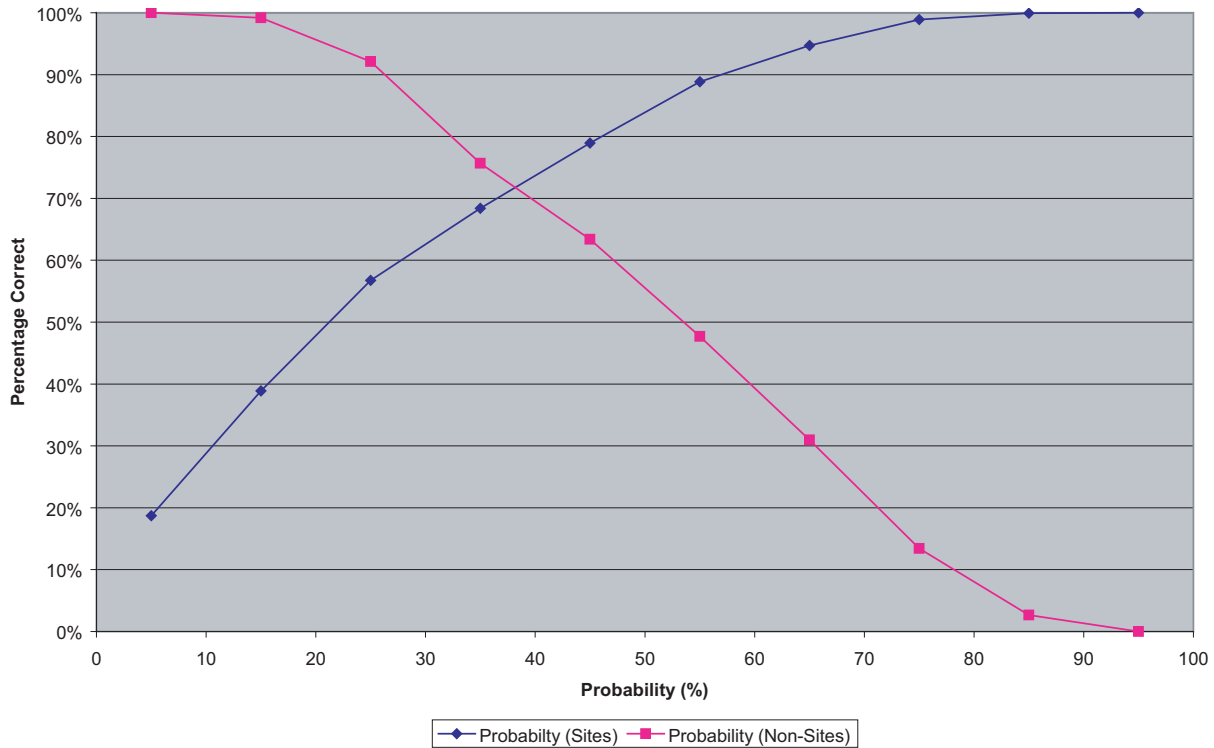


Figure 17. Optimal Cut-Point of Predictive Model (testing data).

Probability levels can be written using the Kvamme's notation as:

$$\Pr(M|S) = 0.8331$$

$$\Pr(M|S') = 0.4905$$

where, $\Pr(M|S)$ is the probability that the model correctly identifies a site given that a site is actually present, and $\Pr(M|S')$ is the probability that the model correctly identifies a non-site given that a site is actually not present (Kvamme 1992:33).

As stated earlier, for a predictive model to be considered successful, the probability of a site occurring given the model specifies a site, $\Pr(S|M)$, must be greater than the base-rate probability $\Pr(S)$ calculated at 0.0335. $\Pr(S|M)$ is the reverse conditional of $\Pr(M|S)$ and can be estimated using Baye's Theorem:

$$\Pr(S|M) = \frac{\Pr(M|S) \Pr(S)}{\Pr(M|S) \Pr(S) + \Pr(M|S') \Pr(S')}$$

Table 2. Results of the testing sample at the 0.25 cut-point level

True Condition	Model Predictions		Testing Data
	M (Site-Present) (L .25)	M' (Site-Absent) (L .25)	N
S (Site-Present)	68543 (83.31%)	13732 (16.69%)	82275
S' (Site-Absent)	4124 (49.05%)	4283 (50.95%)	8407

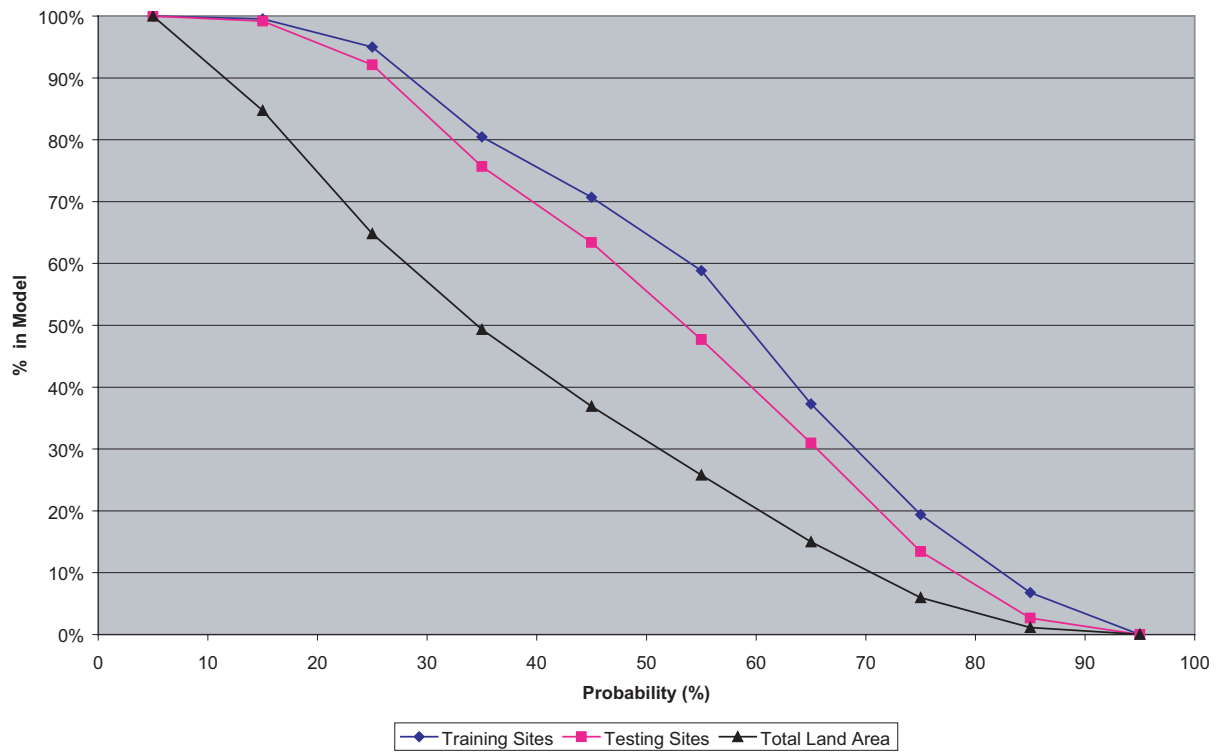


Figure 18. Performance Curves.

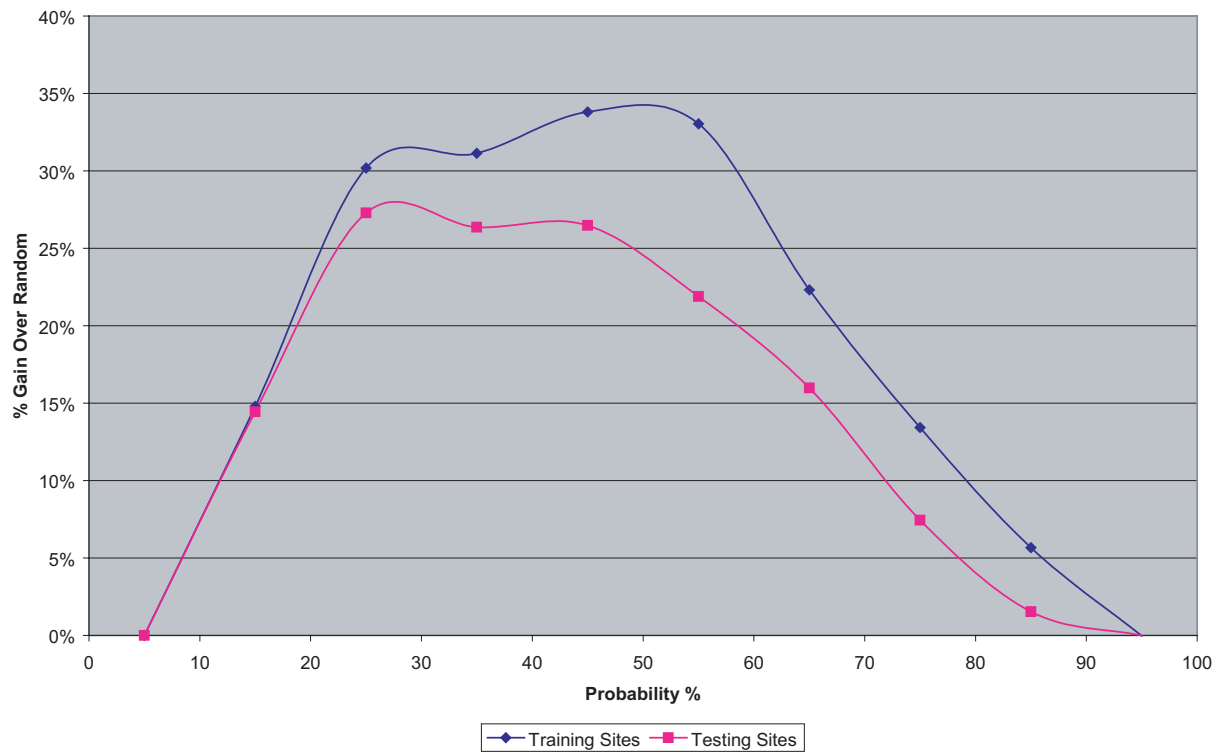


Figure 19. Percent Gain Over Random Classification.

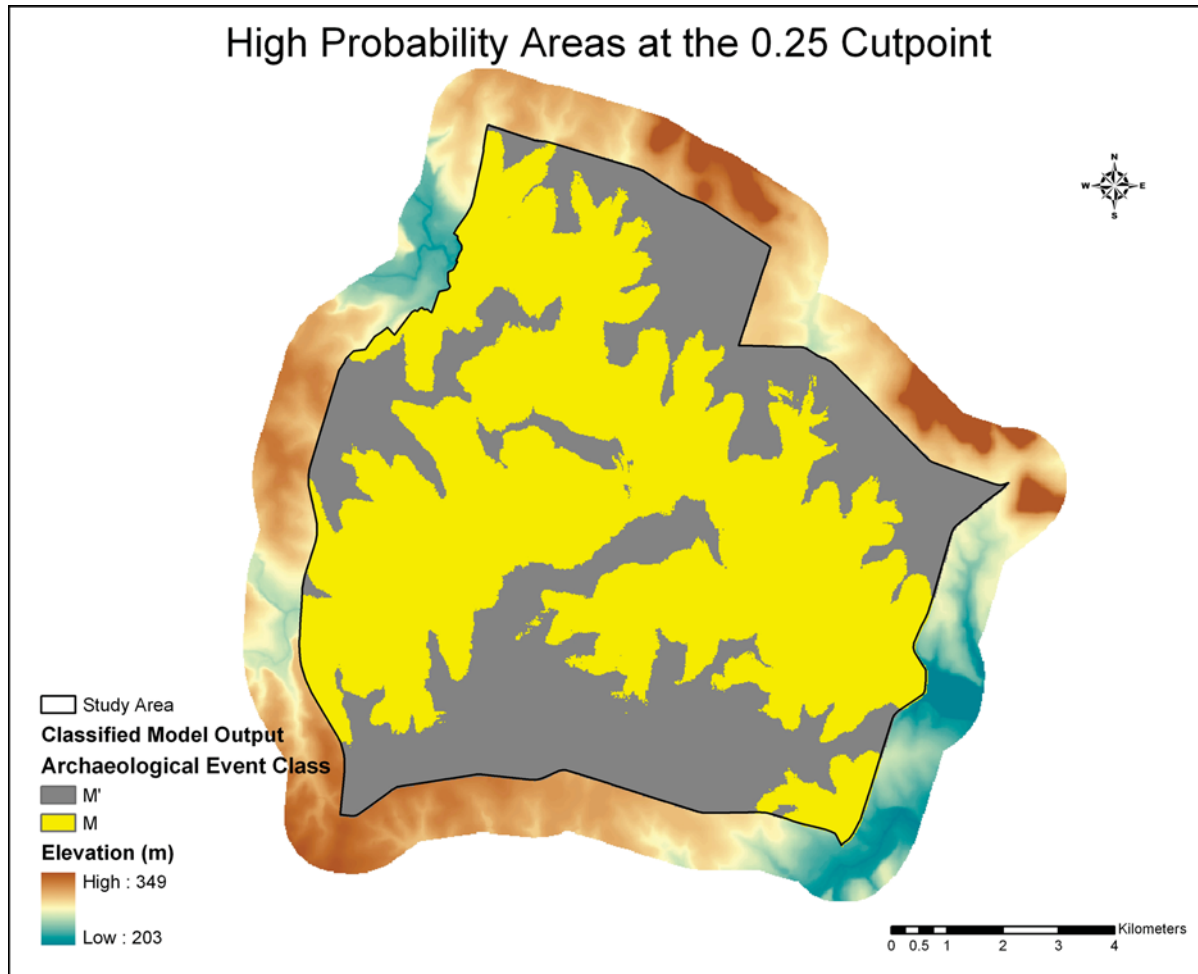


Figure 20. High Probability Areas at the 0.25 Cut-point.

Using the values already determined, this equation yields:

$$\Pr(S | M) = \frac{(.8331)(.0335)}{(.8331)(.0335) + (.4905)(.9665)}$$

$$\Pr(S | M) = 0.0556$$

$$\Pr(S | M) > \Pr(S)$$

$$0.0556 > 0.0335$$

Analysis indicates the probability of a site occurring given that the model predicts a site is = 0.0556. Because $\Pr(S | M)$ is greater than $\Pr(S)$, the current model is more effective than

random chance. Although $\Pr(S | M)$ is very low, it is due to the low base-rate probability and the fact that archaeological sites are rare on the landscape. Approximately 5.6% of the land parcels in the site-present area (M) will contain a site. Stated another way, if the model predicts a site, the probability of a site occurring is $\Pr(S | M) / \Pr(S)$ or $(.0556)/(.0335) = 1.65$ times more likely than random chance alone. Considering that over 4 million land parcels are in the study area, this represents a significant gain over a random chance model (Kvamme 1992).

Using this same methodology, it is possible to estimate the probability of a site occurring given that the model predicts a non-site or $\Pr(S | M')$. Small changes to the above equation result in:

$$\Pr(S | M') = \frac{\Pr(M' | S) \Pr(S)}{\Pr(M' | S) \Pr(S) + \Pr(M' | S') \Pr(S')}$$

Using the values already determined, this equation yields:

$$\Pr(S | M') = \frac{(.132)(.0335)}{(.132)(.0335) + (.5905)(.9665)}$$

$$\Pr(S | M') = 0.0088$$

Calculation of $\Pr(S | M')$ indicates 0.88% of the land parcels in the site-absent area (M') will contain a site. The probability of finding a site is approximately 6 times less than in the site-present area (M). Compared to the base-rate probability, the probability of finding a site is 0.26 as likely as pure-chance. These values represent a significant decrease from the base-rate probability and indicate the model is effectively classifying the landscape into site-present and site-absent classes.

Kvamme (1992) reports similar accuracy values for a model in Pinon Canyon, Colorado. The D-E study area model predicted 85% of sites within 61% of the land area. While more powerful models were created for other zones of the Pinon Canyon study area, Kvamme reports the D-E area values to be a significant improvement over random chance. The most powerful model from Pinon Canyon is reported in Kvamme (1988a). The Pinon Canyon model for the A-B-C study area was able to predict 85% of sites in 39% of the land area, versus the 85% predicted in 56% of the land area for the current model.

5.1.3 2-D Model Conclusions

Summarizing the results of the 2-D model assessment at the 0.25 cut-point level, 85% of site-present land parcels (S) are correctly classified as site-present (M), and 51% of the site-absent land parcels are correctly classified as site-absent (M'). Total area of the study area included in the site-present event class (M) at the 0.25 cut-point level is 56%. In comparison to the base-rate or random chance probabilities, the current model is 1.65 times more likely to predict a site in the site-present class (M) than

random chance alone. Therefore, the model can be viewed as predicting site locations better than random ($0.0556 > 0.0335$), a pre-requisite for any effective model. Archaeological sites should occupy 5.5% of the land parcels within the site-present class (M) and 0.88% of the land parcels in the site-absent class (M'). The base-rate probability indicates a site occurs in 3.35% of the land parcels in the study area.

Model training utilized a set of archaeological sites eligible for listing on the National Register of Historic Places (NRHP) and may represent sites that contain better preserved cultural material than the NRHP non-eligible sites used to test the model. Subsequently, areas with high probability scores may be predicting the location of not only archaeological sites but locations with a high probability of containing well-preserved and extensive cultural materials. However, if the presence of NRHP eligible sites is due to the lack of destruction originating from tracked vehicles, high probability zones may simply reflect areas that are not heavily utilized for training maneuvers. In either case, the implications of using NRHP eligible sites for model training is currently not well understood and requires further investigation.

Disproportionately large values of the known site samples in the 21-30% probability class represent a deficiency in the model. Visual analysis of site locations compared with the model output indicates a preference for site locations at the headwaters of secondary streams, which are consistently modeled as low probability. Refinements in future models should address this qualitative pattern by creating quantitative environmental variables to capture this landscape preference. Assigning the cut-point at the relatively low value of 0.25 can be directly attributed to this pattern.

The current 2-D model differs from the previously created predictive model for Fort Hood (Carlson et al. 1994) in very significant ways, including the land parcel size ($5m^2$ versus $50m^2$), the resolution of the environmental data, the selection of relevant environmental variables, the number of archaeological sites available, and the assessment of model performance. The previous model also included the entire base as the study area. As a result of all these factors, the output of the model (Carlson 1994) differs significantly from the current model. Considering the data quality and extent of archaeo-

logical survey available to the Carlson model, the current model is methodologically superior. However, the only way to ensure the reliability of any predictive model is the acquisition of new archaeological survey data designed for model assessment.

5.2 Geoarchaeological (3-D) Model

One component of the landscape that experiences relatively rapid rates of landform changes includes alluvial systems. The nature of alluvial processes is such that the sub-surface composition of landforms cannot be accurately predicted on the basis of surface morphology. In the absence of high-resolution, sub-surface data (cores, trenches, or exposures), no method can accurately predict the nature and extent of sub-surface alluvial deposits. It is, therefore, beyond the scope of this report to attempt prediction of actual sub-surface stratigraphic units. Previous research into the nature and distribution of Fort Hood alluvial fills and their implications for cultural material preservation represents a reliable source of information for the geoarchaeological component of the predictive model (Nordt 1992). Alluvial landform maps of Cowhouse Creek and Table Rock Creek published in Nordt (1992) form the core of the 3-D model. Alluvial landforms were grouped into T0, T1, and T2 terraces for both stream systems. The landform maps were modified slightly from the original versions for use in this project. The lack of digital versions of the maps required an analog to digital conversion for use in the GIS. Creation of the new landform maps is outlined in Section 4.3 Environmental Database.

Other than the landform maps, Nordt (1992: Table 2) provides specific details concerning the age and cultural potential of alluvial sediments in both stream systems. Information from this table has been reproduced and is shown in Table 3 of this volume. Specifically, Table 3 contains data about the age of landform surfaces, the age range of cultural materials found on landform surfaces, and the potential stratigraphic units that occur within specific landforms. Ages of the stratigraphic units were established using radiocarbon dating. From the chronological data, appropriate cultural periods were assigned to the various stratigraphic units as well as a rating of the potential pres-

ervation of material within the units. Implementation of this data in a GIS environment is discussed in Section 6.0 Model Integration.

Details concerning the sedimentary nature and cultural significance of the alluvial stratigraphic units are beyond the scope of the current report. Detailed descriptions of methodologies, data, and conclusions of the 'Archaeological Geology' model can be found in Nordt (1992).

6.0 MODEL INTEGRATION

Integration of the 2-D and 3-D models required the development of a custom software tool for use within the GIS environment. Development of the 'Predictive Model Viewer' utilized the Visual Basic for Applications (VBA) development environment inside ArcGIS. Microsoft's Visual Basic programming language is used in conjunction with ESRI's ArcObjects software in the VBA environment. ArcObjects is the product name for the software code on which the ArcGIS programs are constructed. Use of ArcObjects in the VBA environment allows customization of the GIS functions available in ArcGIS.

The 'Predictive Model Viewer' allows a user to select any location within the study area and receive information about the location including the UTM coordinate, alluvial landform, stream system, and output of the 2-D and 3-D model. Results of the 2-D model are read directly from the GIS data layer and converted into percentages to simplify interpretation. 3-D model results are based upon Table 2 of Nordt (1992). Predictions about the age and potential preservation of cultural material were extracted and hard coded into the Predictive Model Viewer software code. Information is delivered to the user via a custom Windows-style dialog box. Figures 21 and 22 display the Predictive Model Viewer dialog form and how the form appears within the GIS environment.

An additional function of the tool is the 'Cross Section Viewer'. Users have the ability to open an image of the *generalized geologic cross section* for either stream system. Nordt (1992) constructed generalized geologic cross section diagrams for both the Cowhouse and Table Rock Creek alluvial systems. Although generalized, these diagrams provide valuable information about the potential stratigraphic

Table 3. Preservation potentials for surface and subsurface cultural sites (- none; + low; ++ medium; +++ high; + or ++ facies dependent)

Surface Site Preservation		Subsurface Site Preservation					
Landform	Cultural Divisions	Allostratigraphy ^d	Cultural Divisions	Streams			
				Leon	Cowhouse	Table Rock	Other
T3 ^a , T2	All	Reserve ^a or Jackson (A, Aa, Ab)	-	-	-	-	-
T1 ^a	Paleoindian (?) and younger	Georgetown (B)	Paleoindian	+	+	+	+ or -
T1a ^b , T1 ^c	Early Archaic and younger	Fort Hood (C)	Early Archaic	++	++	+ or ++	+
T1b ^b , T1 ^c	Middle Archaic and younger	West Range (D)	Middle and Late Archaic, Neolithic	+++	+++	++	+
T0	Neolithic and Historic	Ford (E)	Neolithic and Historic	+	+	+	+

^a applies to Leon River only

^b applies to all streams except Cowhouse Creek and the Leon River

^c applies to Cowhouse Creek only

^d applies to all streams unless otherwise stated

Adapted from Nordt (1992)

composition of alluvial landforms. Original figures were scanned and digitized in a graphics program and the layout changed for use in the Cross Section Viewer. The graphic for Table Rock Creek has been adapted from its original content. Originally, the Table Rock Creek graphic, Figure 9 in Nordt (1992), displays the generalized geologic cross section, but does not contain symbology reflective of its cultural significance. Copying the symbology present on the original Cowhouse Creek figure (Figure 28 in Nordt) and incorporating additional information about the cultural significance of

the Table Rock Creek sediments found in Table 2 of Nordt (1992), the appropriate symbology for cultural significance was developed for the new Table Rock Creek graphic. Figure 23a and Figure 23b display the revised geologic cross-section graphics for Cowhouse Creek and Table Rock Creek, respectively.

It is important to note the code for the Predictive Model Viewer will not function if certain map data layers are not within the ArcMap document (see Appendix B for detailed information on the 'Predictive Model Viewer').

Temporal Predictive Model for Fort Hood, Texas

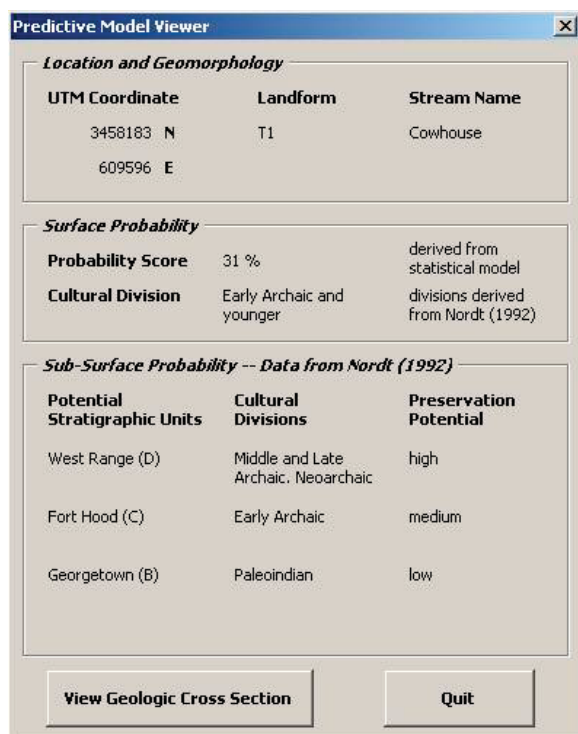


Figure 21. Screen shot of Predictive Model Viewer Graphical User Interface (GUI).

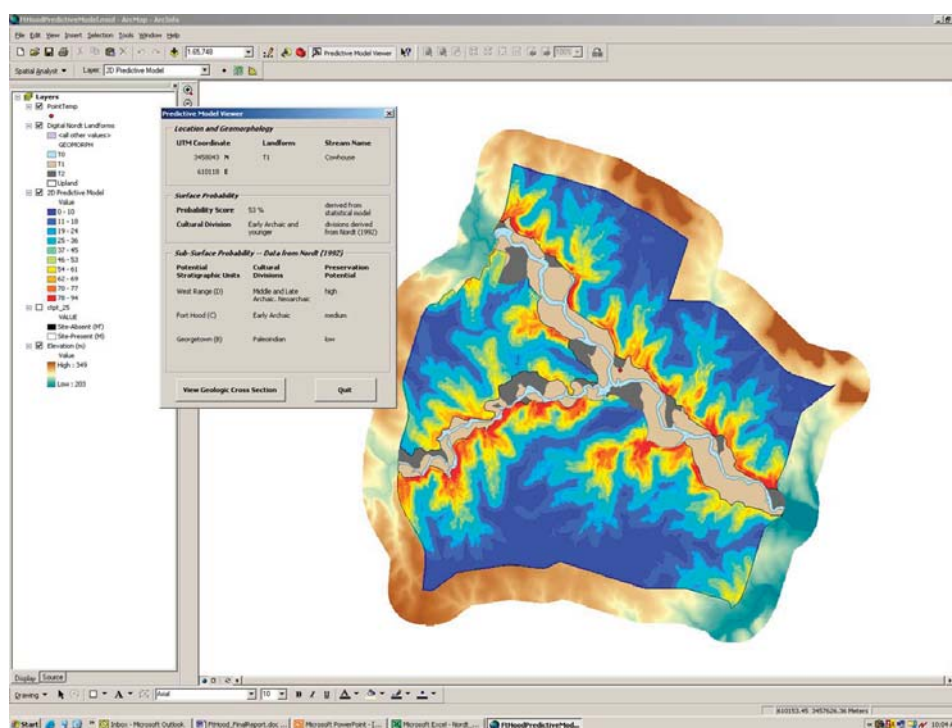


Figure 22. Screen shot of Predictive Model Viewer in operation.

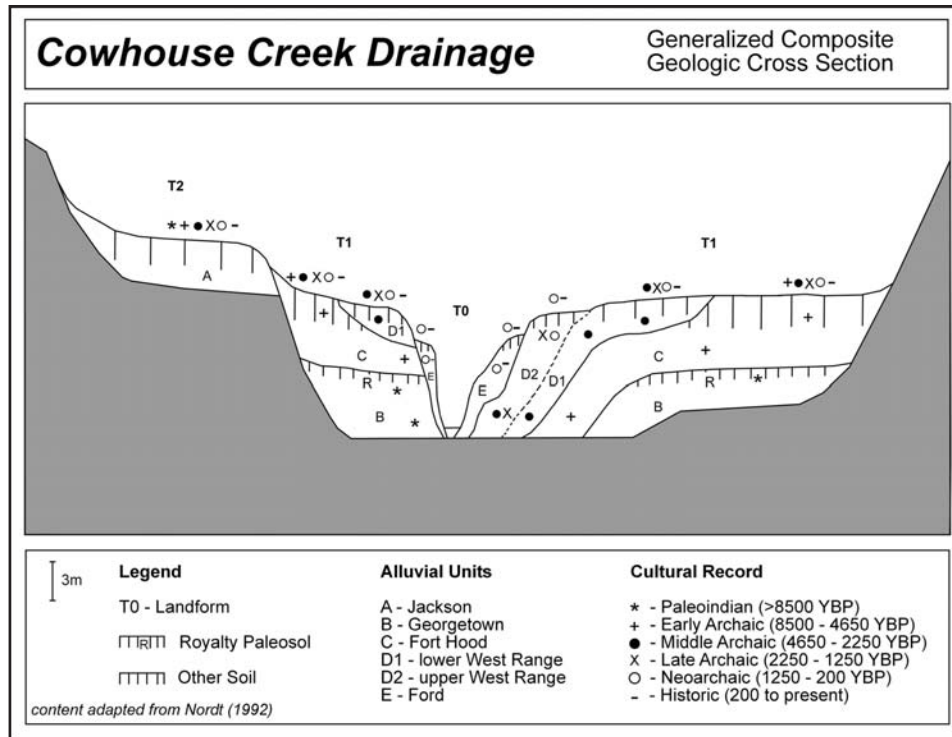


Figure 23a. Cross-section of Cowhouse Creek drainage.

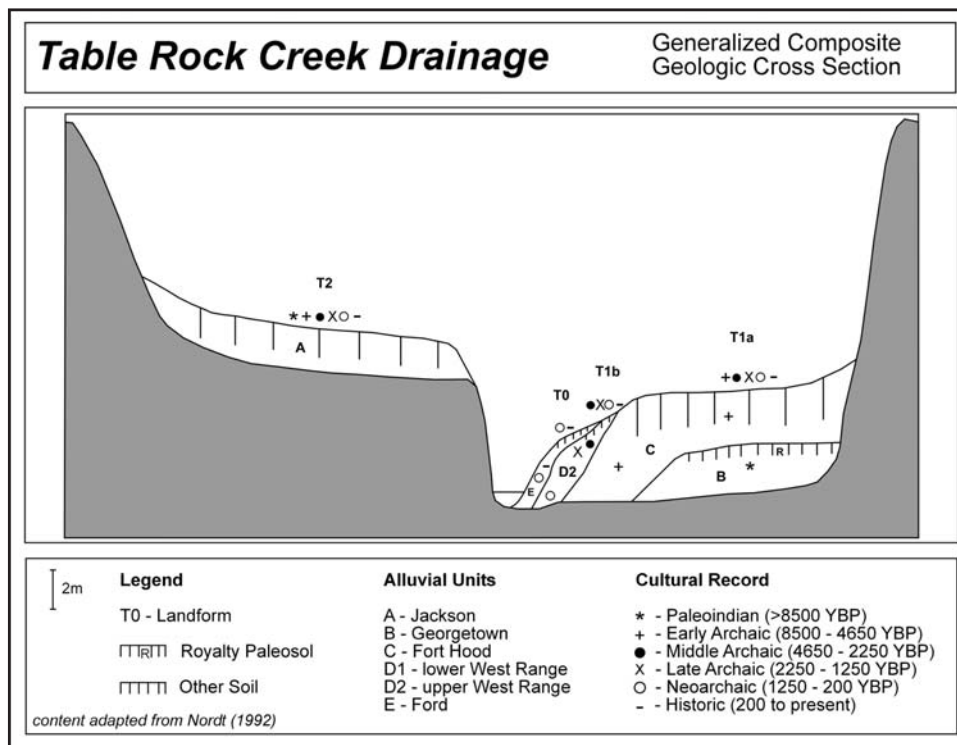


Figure 23b. Cross-section of Table Rock Creek drainage.

7.0 CONCLUSIONS

The methodology used to create this model represents a unique and novel approach to the production and integration of 2-D statistical surface models and sub-surface data in the form of geoarchaeological or 3-D models. Development of the ‘Predictive Model Viewer’ provides a rapid, easy-to-use method of synthesizing the output of the two models.

7.1 Expanding the Methodology

The methodology used to generate the models discussed herein is applicable to the remainder of the Fort Hood Military Reservation. It is the opinion of the authors that modeling the

installation as a whole would increase the predictive power of the 2-D model and provide an empirical equation that could be applied to other areas of the Edwards Plateau. Prerequisites for model construction, including high-resolution digital environmental data, extensive archaeological survey data, and alluvial landform data, are available for the entire base. Geoarchaeological conclusions of other major stream systems reported in Nordt (1992) could be converted to a GIS format using the methodology described herein. The use of a geomorphology layer for the areas outside of the river valleys would likely increase the predictive power of the statistical model and, when combined with sub-surface data for the newly mapped areas, the 3-D model would be truly integrated.

REFERENCES CITED

- Bamforth, D. B.
1988 *Ecology and Human Organization of the Great Plains*. Interdisciplinary Contributions to Archaeology. Plenum Press, New York.
- Barnes, V. E.
1979 *Geologic Atlas of Texas: The Sequin Sheet*. Bureau of Economic Geology, University of Texas, Austin.
- Blum, M. D.
1990 Climatic and Eustatic Controls on Gulf Coastal Plain Fluvial Sedimentation: An Example from the Late Quaternary of the Colorado River, Texas. In *Sequence Stratigraphy as an Exploration Tool*, edited by J.M. Armentrout and B.F. Perkins. GCSSEPMF, Eleventh Annual Research Conference Program and Abstracts.
- Blum, M. D., R. S. Toomey III, and Salvatore Valastro, Jr.
1994 Fluvial Response to Late Quaternary Climatic and Environmental Change, Edwards Plateau, Texas. *Palaeogeography, Palaeoclimatology, Palaeoecology* 108:1-21.
- Blum, M. D., and Salvatore Valastro, Jr.
1989 Response of the Pedernales River of Central Texas to Late Holocene Climatic Change. *Annals of the Association of American Geographers* 79(3): 435-456.
- Butzer, K. W.
1982 *Archaeology as Human Ecology: Method and Theory for a Contextual Approach*. Cambridge University Press, Cambridge.
- Carlson, D. L., J. E. Dockall, and B. W. Olive
1994 *Archaeological Survey at Fort Hood, Texas Fiscal Year 1990: The Northeastern Perimeter Area*. Archaeological Resource Management Series, Research Report No. 24. United States Army, Fort Hood, Texas.
- Carr, C. (editor)
1985 *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory*. Westport Publishers, Kansas City.
- Clark, W. A. V., and P. L. Hosking
1986 *Statistical Methods for Geographers*. John Wiley & Sons, New York.
- Clarke, K. C., B. O. Parks and M. P. Crane
2002 Geographic Information Systems and Environmental Modeling. In *Prentice Hall Series in Geographic Information Science*, edited by K. C. Clarke. Prentice Hall Inc., Upper Saddle River, New Jersey.
- Ebert, J. I.
2000 The State of the Art in "Inductive" Predictive Modeling: Seven Big Mistakes (and Lots of Smaller Ones). In *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit*, edited by K.L. Wescott, and R.J. Brandon, pp. 129-134. Taylor & Francis, London.
- Gaffney, V., and P. M. van Leusen
1995 Postscript - GIS, Environmental Determinism and Archaeology: A Parallel Text. In *Archaeology and Geographical Information Systems*, edited by G. Lock and Z. Stancic, pp. 367-382. Taylor & Francis, London.
- Gibbon, G.
2000 *Appendix A: Archaeological Predictive Modeling: An Overview*. Minnesota Department of Transportation, SHPO Reference Number 95-4098. Minneapolis.
- Hayward, O. T., P. M. Allen, and D. L. Amsbury
1990 *The Lampasas Cut Plain: Evidence for the Cyclic Evolution of a Regional Landscape, Central Texas*. Dallas Geological Society, Dallas.
- Hudak, G. J., E. Hobbs, A. Brooks, C. A. Sersland and C. Phillips
2000 *A Predictive Model of Precontact Archaeological Site Location for the State of Minnesota*. Minnesota Department of Transportation, SHPO Reference Number 95-4098. Minneapolis.
- Jochim, M. A.
1976 *Hunter-Gather Subsistence and Settlement: A Predictive Model*. Studies in Archaeology, Academic Press, New York.

- Kvamme, K. L.
 - 1988a Development and Testing of Quantitative Models. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. J. Judge and L. Sebastian, pp. 325-428. US Department of the Interior, Bureau of Land Management Service Center, Denver.
 - 1988b Using Existing Archaeological Survey Data for Model Building. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. J. Judge and L. Sebastian, pp. 301-323. US Department of the Interior, Bureau of Land Management Service Center, Denver.
 - 1990 The Fundamental Principles and Practice of Predictive Archaeological Modeling. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by A. Voorrips, pp. 257-295. Studies in Modern Archaeology. Holos, Bonn.
 - 1992 A Predictive Site Model on the High Plains: An Example with an Independent Test. *Plains Anthropologist* 37(138):19-40.
- Kvamme, K. L., and T. A. Kohler
 - 1988 Geographic Information Systems: Technical Aids for Data Collection, Analysis, and Display. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. J. Judge and L. Sebastian, pp. 493-547. US Department of the Interior, Bureau of Land Management Service Center, Denver.
- Lock, G.
 - 2000 *Beyond the Map: Archaeology and Spatial Technologies*. NATO Science Series, Series A: Life Sciences. IOS Press, Amsterdam.
- Lock, G., and Z. Stancic
 - 1995 *Archaeology and Geographical Information Systems*. Taylor & Francis, London.
- Nordt, L. C.
 - 1992 *Archaeological Geology of the Fort Hood Military Reservation Ft. Hood, Texas*. Archaeological Resource Management Series, Research Report No. 25. United States Army, Fort Hood, Texas.
- Parker, S.
 - 1985 Predictive Modeling of Site Settlement Systems Using Multivariate Logistics. In *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique and Theory*, edited by C. Carr, pp. 173-207. Westport Publishers, Kansas City.
- Pilgrim, T.
 - 1987 *Predicting Archaeological Sites from Environmental Variables*. BAR International Series 320:133.
- Toomey, R. S., M. D. Blum and S. Valastro, Jr.
 - 1993 Late Quaternary Climates and Environments of the Edwards Plateau, Texas. *Global and Planetary Change* 7: 299-320.
- Verhagen, P.
 - 2000 Archaeology, GIS and Cultural Resource Management. In *Beyond the Map: Archaeology and Spatial Technologies*, edited by G. Lock. NATO Science Series, Series A: Life Sciences. IOS Press, Amsterdam.
- Warren, R. E.
 - 1990 Predictive Modeling in Archaeology: A Primer. In *Interpreting Space: GIS and Archaeology: Applications of Geographic Information Systems*, edited by K. M. S. Allen, S. W. Green and E. B. W. Zubrow, pp. 90-111. Taylor & Francis, London.
- Warren, R. E., and D. L. Asch
 - 2000 A Predictive Model of Archaeological Site Location in the Eastern Prairie Peninsula. In *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit*, edited by K. L. Wescott and R. J. Brandon, pp. 5-32. Taylor & Fisher, London.
- Westcott, K. L., and R. J. Brandon (editors)
 - 2000 *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit*, Taylor & Francis, London.
- Wheatley, D., and M. Gillings
 - 2002 *Spatial Technology and Archaeology: The Archaeological Applications of GIS*. Taylor and Francis, London.

APPENDIX A: GIS DATA LAYERS

The following section lists the GIS layers created for the predictive model. Several of the primary data layers were provided by CEMML. ‘Primary’ data layers were used to derive several ‘secondary’ data layers. Secondary data layers were used in the predictive model. Data layers are categorized into Archaeological, Elevation, Hydrological, Geomorphic, and Predictive Model Output groups. For each data layer, the name is provided as well as the method used to derive it. All raster data layers use a 5 meter² grid cell size.

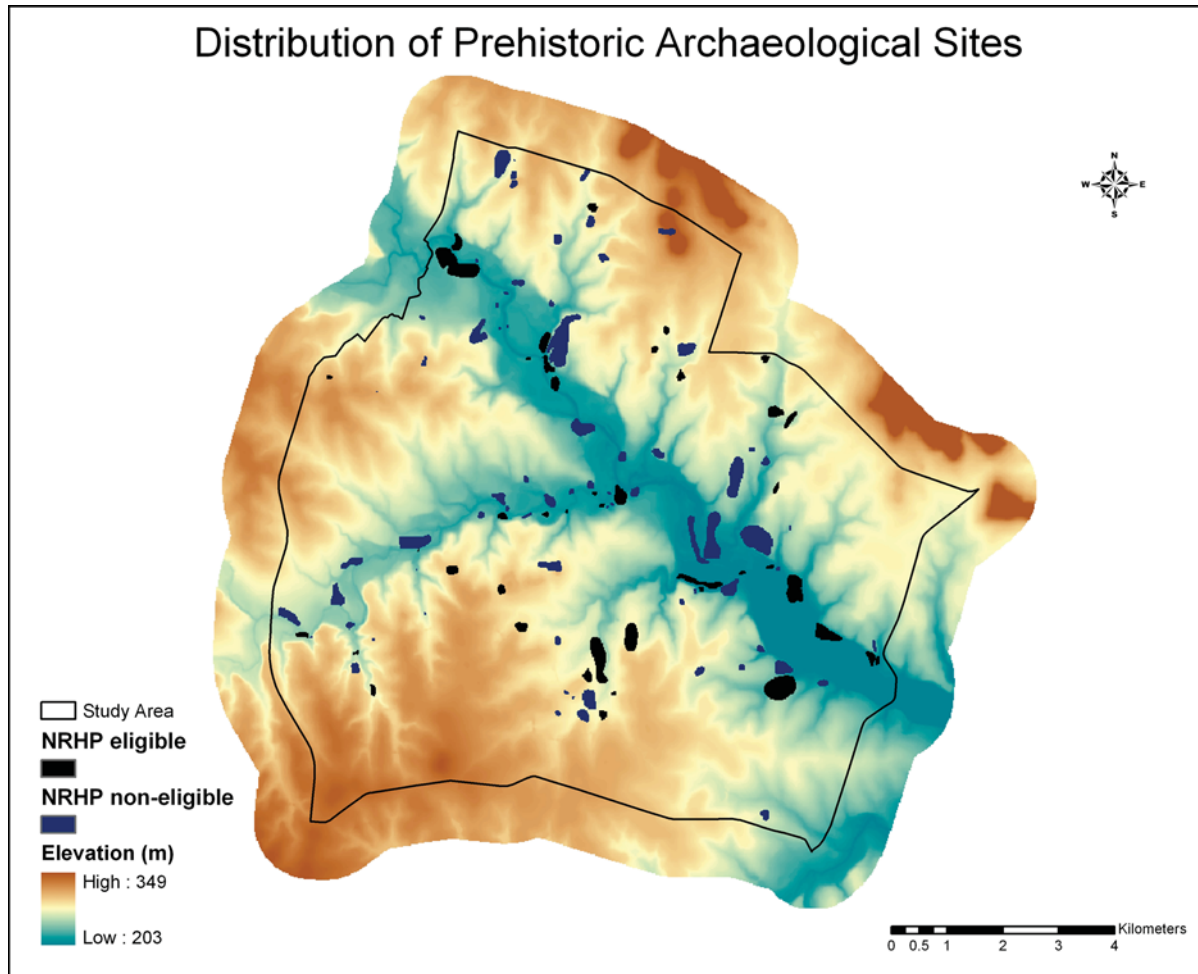


Figure 24. Distribution of Prehistoric Archaeological Sites.

Archaeological Data Layers

- sites_p_elig
 - Historic Register eligible sites extracted from 'hood_s_p' coverage provided by CEMML
 - Contains 44 sites
 - Raster version: sites_p_elig
- sites_p_ne
 - Historic Register non-eligible sites extracted from 'hood_s_p' coverage provided by CEMML
 - Contains 74 sites
 - Raster version: sites_p_ne

- See Figure 24 for the distribution map of all archaeological sites.

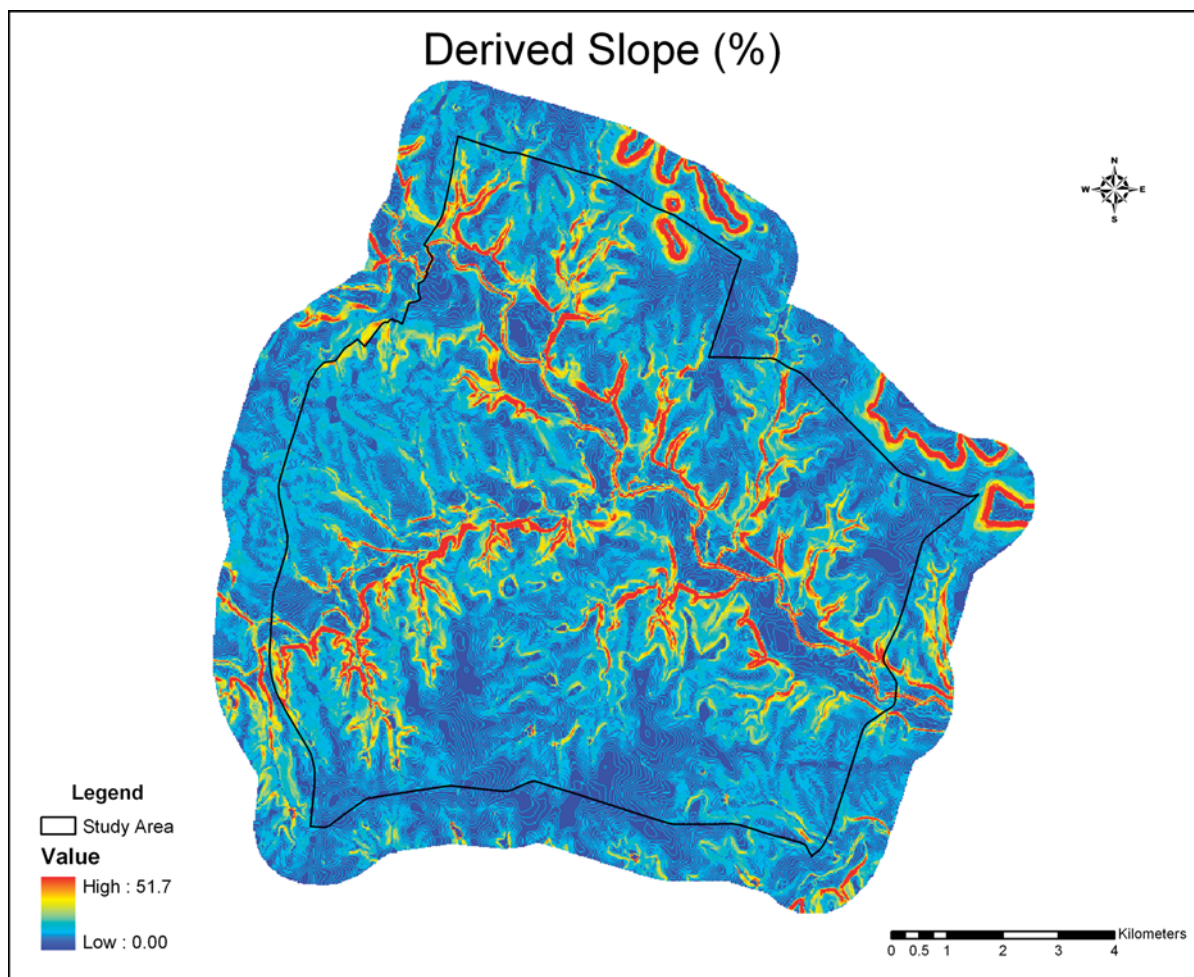


Figure 25. Derived Slope.

Environmental Data Layers

Elevation: 5m grid cell (very high resolution)

- dem_int
 - INT (lfhypdem)
 - Source data provided by CEMML: DEM is an interpolation of 3-meter contour map provided by Ft. Hood, exact contour interval unknown (See 'lfhypdem' metadata for details)
 - See Figure 3 (on page 4)
- slope
 - Spatial Analyst (dem_int)
 - See Figure 25

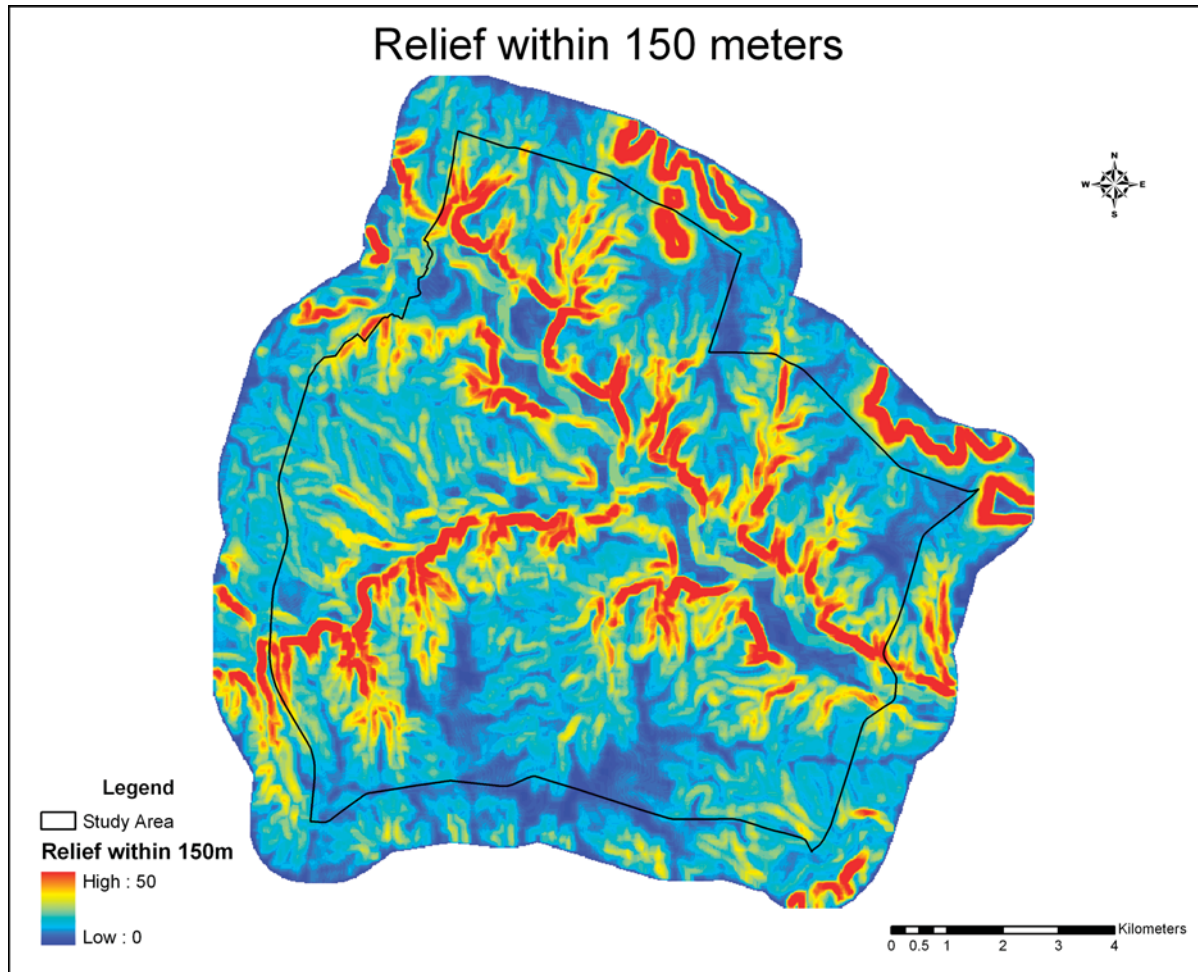


Figure 26. Relief within 150 meters.

- relief150
 - Focalrange (dem_int, circle, 30)
 - See Figure 26

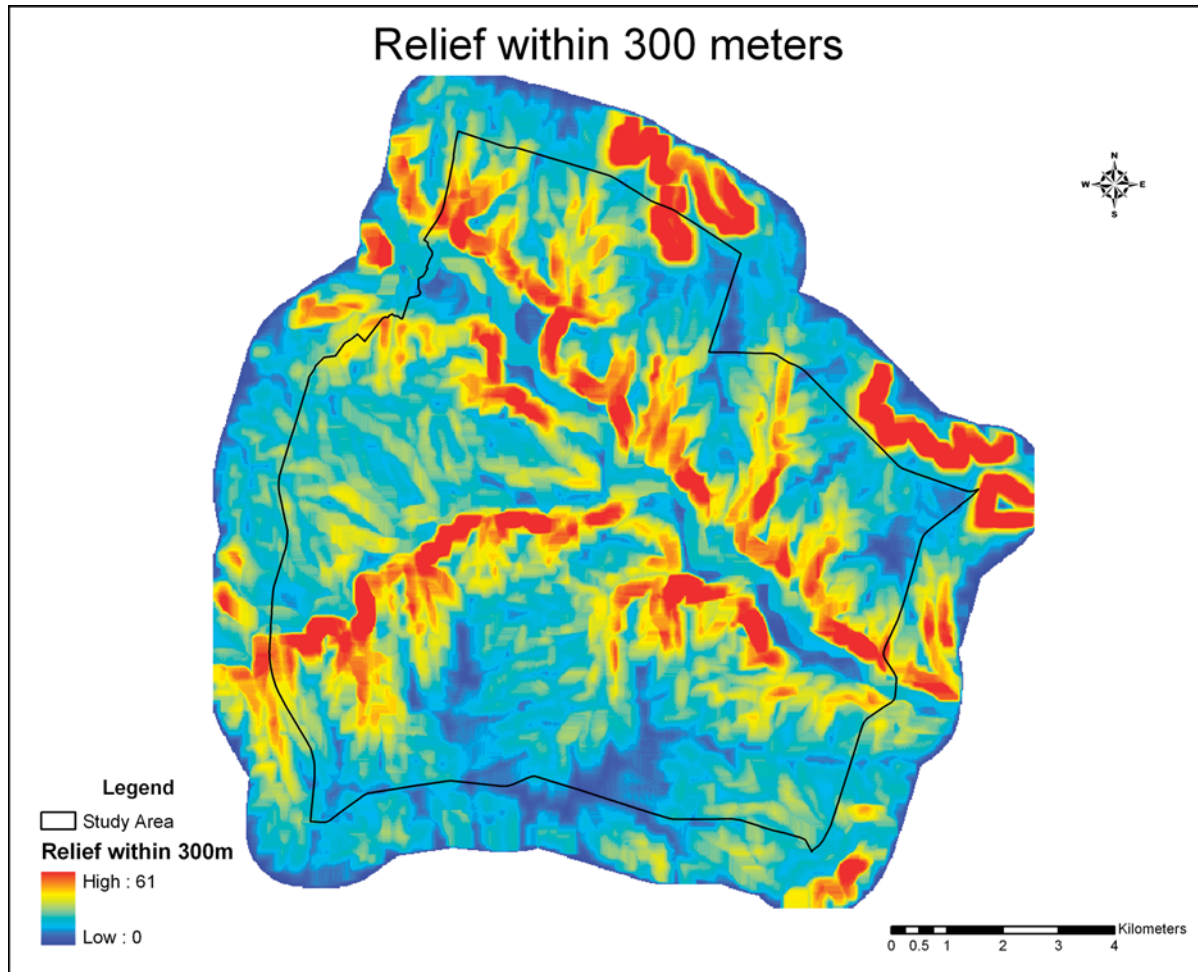


Figure 27. Relief within 300 meters

- relief300
 - Focalrange (dem_int, circle, 60)
 - See Figure 27

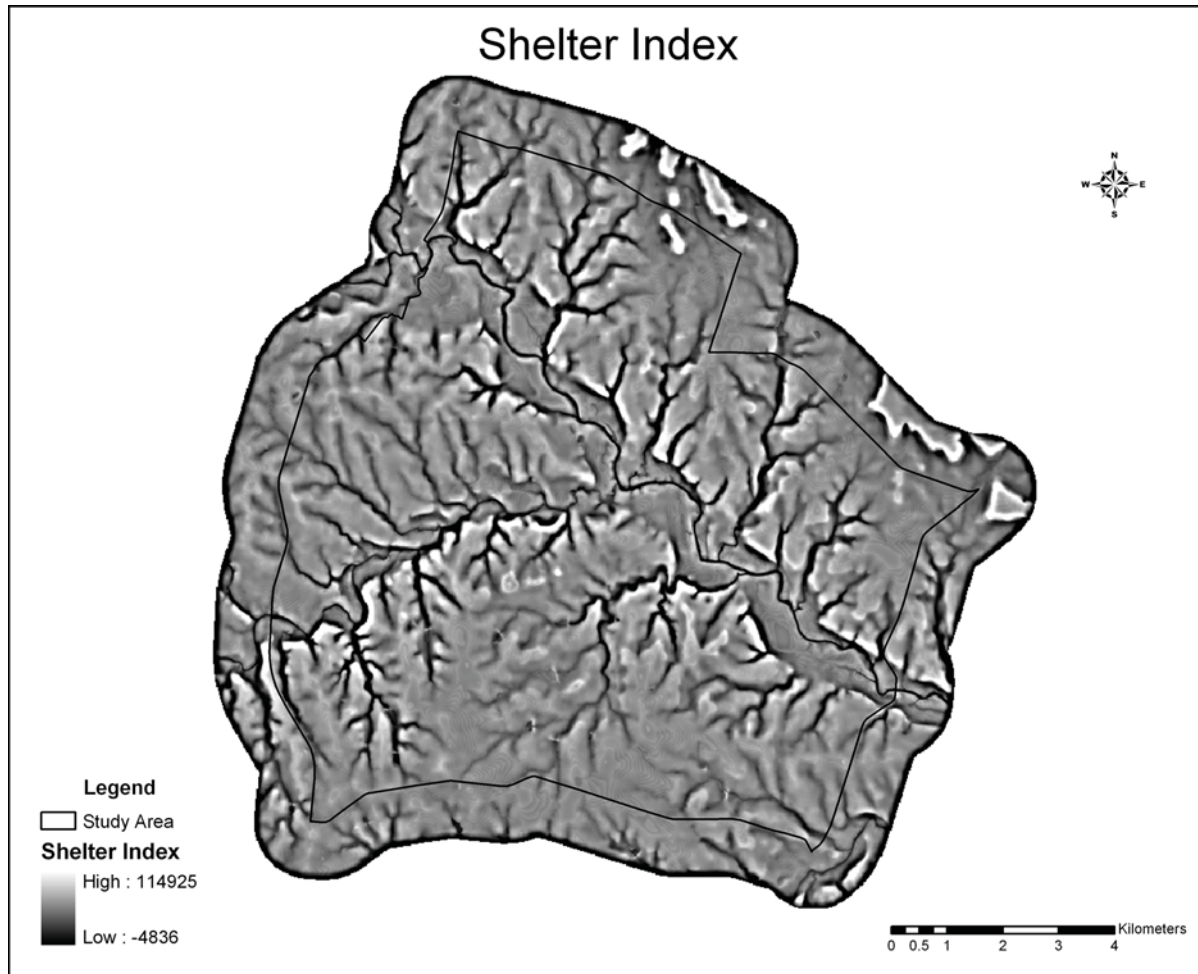


Figure 28. Shelter Index.

- shelter150 (Shelter Index)
 - temp1 = focalsum ([study_area], circle, 30, data)
 - temp2 = ([dem_int]+20) * [temp1]
 - temp3 = focalsum ([dem_int], circle, 30, data)
 - shelter150m = [temp2] - [temp3]
 - See Figure 28

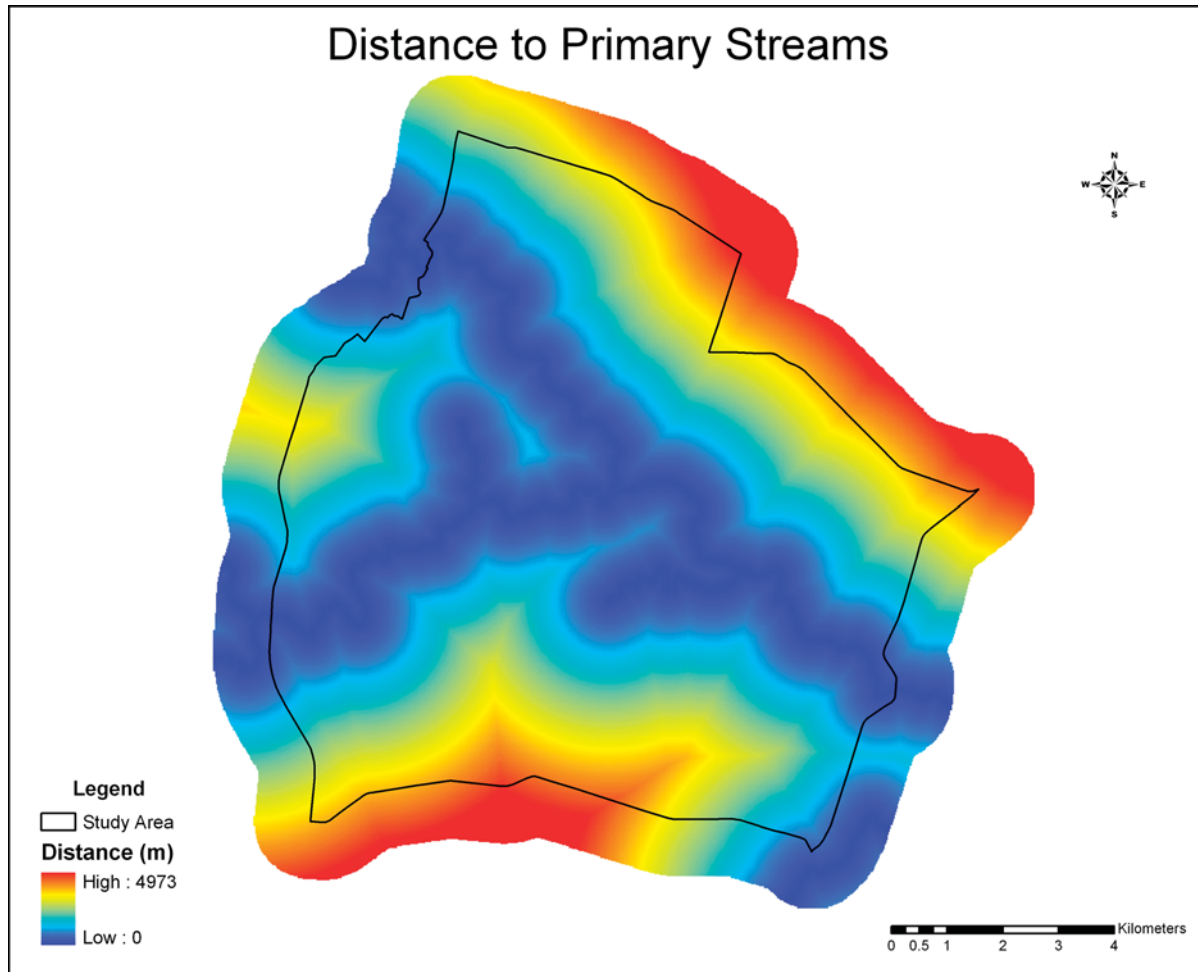


Figure 29. Distance to Primary Streams.

Hydrologic: 5m grid cell (very high resolution)

- streams_main
 - ‘main’ streams extracted from ‘hysurwcc’ coverage, provided by CEMML, using large field map created for Jim Zeidler (10/08/2002)
- streams_secondary
 - ‘secondary’ streams extracted from ‘hysurwcc’ coverage, provided by CEMML, using large field map created for Jim Zeidler (10/08/2002)
- d_main
 - Spatial Analyst: Distance from..... Straight line (streams_main)
 - See Figure 29

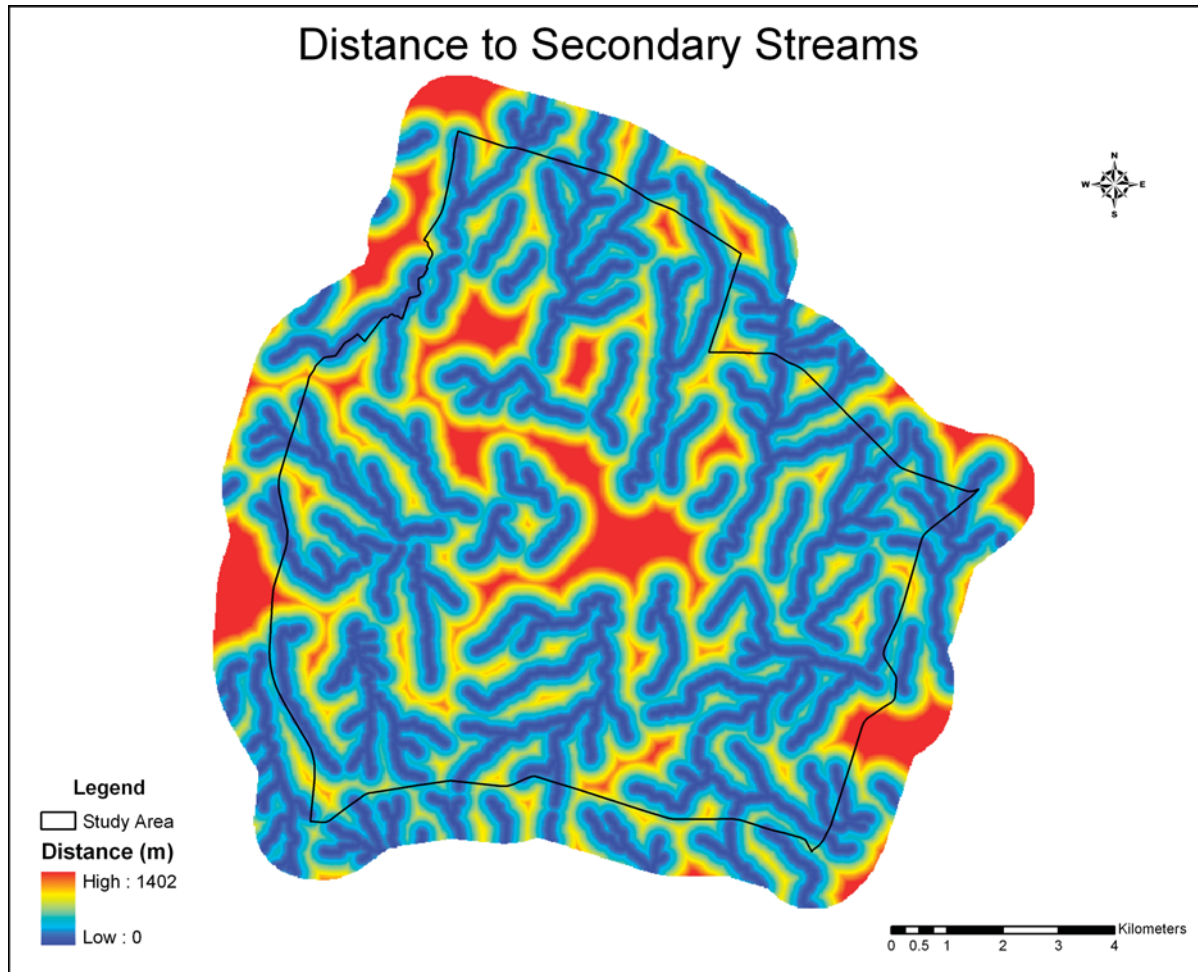


Figure 30. Distance to Secondary Streams.

- d_secondary
 - Spatial Analyst: Distance from.....
Straight line (streams_secondary)
 - See Figure 30

Geomorphological / Landscape

- Digital Nordt Landforms
 - Polygon shapefile created by re-mapping the alluvial landforms of Cowhouse and Table Rock Creek
 - The published landform maps in Nordt (1992) were used as a guide
 - Remapping took place within the ArcMap GIS environment with a digital elevation model, derived slope data, aerial imagery, and 1:24k topographic maps as additional information
 - See Figure 8 (on page 14)

Predictive Model Data Layers

- stathoodmodel – referred to as ‘2D Predictive Model’ in the ArcMap document
 - created using the logistic regression equation described in the body of the report
 - utilized in the Predictive Model Viewer
 - each grid cell contains a probability percentage (0-100%) which was created by multiplying the original probability model by 100
 - See Figure 9 (on page 17)
- ctp25
 - probability map reclassification using the 0.25 cut-point value into site-absent (M') and site-present (M) event classes
 - serves as the basis of model accuracy assessment
 - See Figure 20 (on page 26)

**APPENDIX B:
PREDICTIVE MODEL VIEWER
(NOT AVAILABLE TO THE GENERAL PUBLIC)**

As described in the main body of the report, the Predictive Model Viewer is a custom GIS tool designed using Visual Basic and ArcObjects. Specifically, the tool allows users to select a location within the study area and receive 2-D and 3-D model predictions about the potential of the location to contain cultural material. Software code used to construct the tool was developed within the Visual Basic for Applications (VBA) environment embedded within ArcGIS. Access to the Visual Basic code is 'locked' within the FtHoodPredictiveModel.mxd ArcMap document. Using the password 'fthood', users can access the code. Locking of the VBA environment was done simply to stop inadvertent code changes.

The 'Predictive Model Viewer' and associated 'Cross-Section Viewer' require certain GIS data files and other graphic (.jpg) files to be located in specific locations relative to each other on a computer hard drive. The folder that contains the ArcMap document and other required files can be placed at any hard drive location. However, in order for the 'Predictive Model Viewer' to operate correctly, the directory structure of the files should not be changed, and the files should be kept together at the same location.

Figure 22 displays a screen shot of the operating Predictive Model Viewer. Note the presence of the required GIS data files within the ArcMap Table of Contents. Specifically, 'Point-Temp', 'Digital Nordt Landforms', and '2D Predictive Model' must be co-located within the ArcMap document for the Predictive Model Viewer to work correctly. The 'Cross-Section Viewer' requires that the three graphics files named 'CowhouseSection', 'TableRockSection', and 'UplandSection' be located within a folder entitled 'graphics' at the same location as the ArcMap document (FtHoodPredictiveModel.mxd). These files are distributed with the correct directory structure on a CD format. Files

must simply be moved, or the .zip file opened, to a single hard drive location. If the ArcMap document becomes corrupt or the 'Predictive Model Viewer' does not operate correctly, the original .mxd file can be installed again from the original CD.

The 'FtHoodPredictiveModel.mxd' file has been write-protected, which means users should not be able to save any changes to the ArcMap document using the original name. It is hoped this will limit inadvertent changes that would adversely affect the 'Predictive Model Viewer' from operating correctly. Other GIS data layers may be added to the ArcMap document and saved. However, the new ArcMap (.mxd) file must be saved using a different name. As long as the three required GIS data layers are within the document, the 'Predictive Model Viewer' tool will operate correctly regardless of the name of the ArcMap document. The 'Cross-Section Viewer' tool will only work if the 'graphics' folder, which contains the cross-section images, is located at the same directory location as the new ArcMap document.

For example, installing the ArcMap document at this location:

C:\FtHood_GIS\Model\
FtHoodPredictiveModel.mxd

requires that the graphics folder be installed at this location:

C:\FtHood_GIS\Model\graphics

Contact Joshua S. Campbell (jsc1@ku.edu) for any additional questions concerning the construction or operation of the 'Predictive Model Viewer' and 'Cross-Section Viewer.' NOTE: This product is not available to the general public and is *not* available on the enclosed CD-ROM disk.